MD281
50¢

The NEW Mathematics

+

IRVING ADLER

Instructor in Mathematics, School of General Studies,
Columbia University

☞ **Teachers, students, and general readers: This is the first book to explain—in simple, uncomplicated language—the working methods, ideas, and fundamental concepts of the revolutionary developments in modern mathematics.**

A MENTOR BOOK

The Wonderful World of Modern Mathematics

In this era of expanding science, modern mathematics is like a jet plane streaking across the sky to an adventurous destination. But to the average bystander who watches the jet-stream, even its t[illegible] is a mysterious, hidden secret. The [illegible] book is to help that bystander pe[illegible] known and fascinating realm of [illegible] imaginary, rational, and complex. [illegible]

The author maps the routes of advance[illegible] their fundamental origins, step by step, [illegible] how our system has grown from the w[illegible] bers we use in counting to measurements i[illegible] that must be denoted by arrows.

Anyone who has a smattering of high-school algebra and geometry can follow the do-it-yourself sections at the end of each chapter and in the supplement. Pencil in hand, he will navigate through old structures into new, and experience the thrill of discovery that awaits him in the field of modern mathematics.

THIS IS A REPRINT OF THE HARDCOVER EDITION
ORIGINALLY PUBLISHED BY THE JOHN DAY COMPANY

r Books of Special Interest

AN *by Irving Adler*
account of what science has discovered
rigin of life. Preface by Linus Pauling.
(#Ks369—35¢)

OUSE OF NUMBERS *by Irving Adler*
ematical curiosities, riddles, tricks, and games that
h the basic principles of arithmetic. (#Ks361—35¢)

HE STARS: Steppingstones into Space *by Irving Adler*
A clear explanation of the mysteries of the heavens, with photographs and line drawings. (#Ks364—35¢)

NEW HANDBOOK OF THE HEAVENS *by Hubert J. Bernhard, Dorothy A. Bennett* and *Hugh S. Rice*
A guide to the stars and planets. (#MD114—50¢)

TO OUR READERS

We welcome your request for our free catalog of SIGNET and MENTOR BOOKS. If your dealer does not have the books you want, you may order them by mail enclosing the list price plus 5¢ a copy to cover mailing. The New American Library of World Literature, Inc., P.O. Box 2310 Grand Central Station, New York 17, New York.

By Irving Adler

THE NEW
MATHEMAT

With diagrams by
Ruth Adler

A MENTOR BOOK

Published by THE NEW AMERICAN LIBRARY

© 1958 by Irving and Ruth Adler
© 1959 by Irving Adler

All rights reserved. This book, or parts thereof, must not be reproduced in any form without permission. For information address The John Day Company, 62 West 45th Street, New York 36, N. Y.

Published as a MENTOR BOOK
By Arrangement with The John Day Company

First Printing, May, 1960
Second Printing, October, 1960

MENTOR BOOKS are published by
The New American Library of World Literature, Inc.
501 Madison Avenue, New York 22, New York

PRINTED IN THE UNITED STATES OF AMERICA

Contents

Foreword

I. *Numbers for Counting*

II. *Number Systems Without "Numbers"*

III. *New Numbers From Old*

IV. *Numbers for Measuring*

V. *Filling Out the Line*

VI. *Spilling Into the Plane* 125

VII. *The Rank and File: Matrices* 148

VIII. *Arrows That Are Numbers* 160

Bibliography 181

Summary of Basic Definitions 182

Do-It-Yourself Supplement 184

Index 191

Foreword

AS WE go about our daily busines[illegible] quent use of whole numbers, like 1, 2, 3, an[illegible] when we count collections of objects. We a[illegible] like ½ and ¼, which arise when measurem[illegible] These are obviously two different kinds of n[illegible] fractions can never arise as a result of coun[illegible] what sense then, do they both deserve to be c[illegible] bers? When we say that the fraction 4/2 is equ[illegible] whole number 2, what does this mean? How can a [illegible] of one kind be equal to a number of a totally different k[illegible]

These are among the questions which modern ma[illegible]matics has explored, and for which it has found answe[illegible] Here are others:

We sometimes have occasion to use a number like the square-root of two. It seems like a very elusive number that is reluctant to show its face. Either it hides shyly behind a symbol like $\sqrt{2}$, or it reveals itself only piecemeal as a decimal, 1.4, 1.41, 1.414, in a process that never ends. Why doesn't it behave itself and settle down like a decent ordinary fraction with a definite numerator and denominator that can determine its value once and for all?

In elementary algebra we were introduced to negative numbers, and taught such mysterious rules as that the product of two negative numbers is a positive number. Where does such a rule come from?

The electrical engineer uses the number $\sqrt{-1}$ in the equations that describe the behavior of an alternating current. This number is called "imaginary," yet there is nothing imaginary about the electrical current it helps to describe. It is a genuine number, mathematicians assure us, although it is not real. What is the meaning of this paradox?

These are some of the questions that we shall look into in the course of this book as we take a close look at the numbers of everyday life. We shall find the answers in the ct that our number system has not been static, but has wing, while our conception of what constitutes a hanged. As we trace this growth, we shall dis- liar roots of the unfamiliar concepts and n mathematics.

tics, one of the oldest of the sciences, is ne vigor and vitality of youth. It is con- ing into new areas of investigation, and w concepts that are the fruit of a century-old in mathematical thinking. Associated with the s is a new vocabulary that gives modern mathe- writing its characteristic flavor. To the mathema- n, the new ideas expressed by the new words serve as ight light that penetrates to the core of a problem and lps him see and understand. To the layman, the new vocabulary is often an opaque screen behind which things are going on that he feels he cannot hope to understand. The purpose of this book is to remove that screen, by introducing the reader to the meaning of some of the basic ideas of modern mathematics.

This book is addressed to the average reader who is curious about the new developments in mathematics. It is not a refresher course in high school mathematics. It is not a rehash of old ideas, but an introduction to new ideas, traditionally presented only to the specialist, in advanced mathematics courses on the college senior or graduate level. But, although the ideas are advanced, the presentation is elementary. Anybody who has had high school algebra and geometry will be able to understand and enjoy this book.

A typical text in advanced mathematics today bristles with such terms as group, ring, field, homomorphism, isomorphism, and homeomorphism. These unfamiliar looking words make it seem as though mathematics has abandoned its old subject matter, and is no longer concerned with the study of numbers and space. This, of course, is not true. Numbers and space are still very much at the heart of mathematics. The new ideas and terms have arisen in connection with a more penetrating analysis of their properties.

Underlying the terms group, ring, and field, for example, are the old, familiar, simple operations of additio[illegible] traction, multiplication, and division. The math[illegible] has discovered that these operations are not [illegible] property of numbers alone. So he studies them [illegible] general form, in order to discover rules that [illegible] in any context in which the operations are pe[illegible]

The outlook of the modern mathematicia[illegible] in his frequent use of the word stem "morp[illegible] form, as in the words homomorphism, isomorp[illegible] homeomorphism. The mathematician sees the num[illegible] tem as a complex of interrelated *structures.* He s[illegible] these structures separately, and in their relationships [illegible] each other. The exploration of these structures has revealed that we have, not *a* number system, but number systems; not algebra, but algebras; not geometry, but geometries; not space, but spaces. While the properties of numbers and space have been generalized, the subject matter of mathematics has been pluralized.

The central thread around which the book is organized is the expansion of the number system, from natural numbers to integers to rational numbers to real numbers to complex numbers. Although this sequence of steps in the development of the number system parallels very roughly historical stages in the development of the concept of number, the organization of the book is not chronological or historical. It is a logical organization from the modern point of view, showing how the various number systems are related to each other. The development outlined here might be referred to as "operation bootstrap." The system of natural numbers (the whole numbers used for counting) has defects that limit its usefulness. The story presented here shows how mathematics has lifted itself by its own bootstraps, using the defective system of natural numbers to construct bigger and better number systems that eliminate the defects.

At each stage of the construction of the expanded number systems, we encounter some of the structures such as groups, rings and fields that receive so much attention in mathematics. These modern concepts are intro- by means of familiar examples in the number then other less familiar examples are given, ds the sections devoted to these modern con- er will be aware of the fact that he is merely corner of a great rug that has a beautiful but n woven into it. If what he sees from the es his curiosity about the main design, it is at he will satisfy this curiosity by systematic om some of the standard text-books. A bibliography en at the end of the book.

To get the most value and enjoyment out of this book, read it with pencil and paper in hand. Verify the steps of each argument, work through all examples given, and do other examples like them. A "Do It Yourself" section at the end of each chapter gives you an opportunity to strengthen through use your understanding of the new ideas you will acquire in the book.

THE NEW MATHEMATIC

Numbers for Counting

WE ARE thoroughly familiar with the faces [illegible] people with whom we live. Yet we are rarely co[illegible] the details of their features. If, as we look at a f[illegible] we do take particular notice of the details, su[illegible] of the lip, or a line in the forehead, it seems [illegible] are seeing them for the first time. Then, seeing [illegible] that we never notice, we suddenly have the feel[illegible] are looking at the face of a stranger. We shall hav[illegible] experience with the familiar numbers of everyday [illegible] we use these numbers, we take advantage of certain [illegible]erties that they have. However, we are so accustomed [illegible] these properties that we are hardly aware of them as we use them. We shall now take particular notice of these properties, and list them explicitly. Looking at the familiar features of ordinary numbers, we shall see the strange new face of modern mathematics.

The first numbers we all learn to use are those we need to answer the question, "How many?" They are the numbers 1, 2, 3, 4, 5, and so on. There is an endless chain of these numbers. We use them for counting, and we perform calculations with them, such as addition and multiplication. Let us take a close look at these simple acts.

Counting

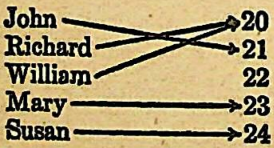
Suppose that on Tuesday evening you want to see how many days are left till the end of the week. It is likely that you will take count in this way: You will call off the names of the days, Wednesday, Thursday, Friday, and Saturday, and, for each day that you name, you will turn down one finger on your right hand. After completing the list of days, you find that you have turned down all the fingers on your right hand except the thumb. So you conclude that there are *four* days left till the end of the week. We find hidden in this procedure three important mathematical concepts: the idea of a *mapping,* the idea of a *one-to-one correspondence,* and the idea of *cardinal number.*

A mapping is a matching operation between two sets of bjects: to each member of one set a member of the other ssigned as partner. The two sets in this case are the ys being counted, and the set of fingers on your et up a mapping when you single out a finger to r each day you count. The mapping might be n the following table:

Wednesday → little finger
Thursday → ring finger
Friday → middle finger
Saturday → index finger

arrowheads indicate that the mapping has a direction. You select a finger for each day you name. This is not the same as selecting a day for each finger. To specify the direction of the mapping, we say that it is a mapping of the set of days named *into* the set of fingers. We refer to the finger into which a day is mapped as its *image* under the mapping.

Another mapping is shown in the diagram below. In this mapping, a set of names of people has been mapped into the set of whole numbers from 20 to 24 by assigning to each name the person's age in years:



This mapping differs from the other one in one important respect. The two names, Richard and William, are both mapped into the same number. This is an example of a *many-to-one mapping*, in which a single object may be the image of more than one object. In the mapping of days into fingers, however, no two days were mapped into the same finger. This is an example of a *one-to-one mapping*, in which each object is the image of at most one object.

In the mapping of the set of days into the fingers of the right hand, one of the fingers, the thumb, wasn't used at all. For this reason the mapping of the set of days into the set of fingers on the right hand is not reversible. If we try to reverse it, we find that there is no mapping of the [illegible] into a day. We do not consider it a mapping th[illegible] a mapping should provide an image for each [illegible] set on which the mapping is performed. [illegible] consider *only the set of fingers turned down*, [illegible] ping is reversible. Then, while each day named [illegible] rate finger as its image, in the reverse mapping, [illegible] has a separate day as its image. In this case we say [illegible] two sets are in *one-to-one correspondence*. Two sets [illegible] one-to-one correspondence when there is a reversible [illegible]ping that assigns each member of one set to one and only one partner in the other. The diagram below, using double-headed arrows, shows the one-to-one correspondence between the set of days and the set of fingers turned down:

Wednesday	⟷	little finger
Thursday	⟷	ring finger
Friday	⟷	middle finger
Saturday	⟷	index finger

When two sets can be put into one-to-one correspondence by some mapping, we say that they contain the same number of objects, or have the same *cardinal number*. All sets that have the same cardinal number can be put into one-to-one correspondence with each other. They form a family of sets associated with that cardinal number. Each cardinal number has its own family of sets. For example, sets consisting of *single objects only* belong to the family of sets associated with the number we call *one*. Sets of *pairs of* *objects* belong to the family of sets associated with the number we call *two*. Sets of triples belong to the family of sets associated with the number we call *three*, and so on.

Any set we ever deal with belongs to one of these families. When we ask the question, "How many objects are there in the set?", it is really like asking, "Which family of sets does it belong to?" To answer the question, we follow thi procedure: We pick one set from each family, and use it a a standard set for making comparisons. We match the se e are interested in against these standard sets, until we with which it can be put into one-to-one corre In this way we identify the family of sets tha o, and the cardinal number associated with tha is precisely what you do when you match day rs. You use the set consisting of your littl s a standard set to represent the number *one* e set consisting of little finger and ring finger as a d set to represent the number *two*. You use the set ing of little finger, ring finger, and middle finger as ndard set to represent the number *three*. The set consisting of little finger, ring finger, middle finger, and index finger is your standard set representing the number *four*. That is why you drew the conclusion, in this case, that there are four more days to the end of the week.

On other occasions, we use a method of counting that is more sophisticated but is essentially the same. We count out four objects by saying to ourselves, "one, two, three, four." As we count, we set up a one-to-one correspondence between the objects we are counting and sets of spoken number-names. The first object is matched with the set consisting of the single word, "one." The first two are matched with the set consisting of the words, "one, two." The first three are matched with the set consisting of the words, "one, two, three." And so on. By using the number names of the numbers in order of size, we keep enlarging the standard set step by step. When the count ends, we know that the last number-name used is the cardinal number o the last standard set against which we matched the objects we are counting. So it is also the cardinal number of the counted objects. By using standard sets made up of number names arranged in order we telescope a whole series of matching operations into one, and end up with the answer to the question, "How many?"

Addition

A typical problem in addition is to find the sum of the numbers 2 and 3. The meaning of this problem can be r[illegible]stated in terms of sets of objects in this way: Su[illegible] have one set of objects whose cardinal numb[illegible] another set of objects, different from those in [illegible] and whose cardinal number is 3. A larger [illegible] when these two sets are united. What is the [illegible] ber of the united set?

We can answer this question by actually for[illegible] united set, and then identifying the standard set [illegible] it can be put into one-to-one correspondence. Th[illegible] procedure of the beginner in arithmetic, who first [illegible] down two fingers, then turns down three more fingers, and finally matches the set of turned down fingers against the standard set consisting of the spoken words, "one, two, three, four, five." However, experienced calculators use a short-cut for getting the answer. Having carried out the process of uniting and counting sets of objects many times before, we record the results in an *addition table* which we memorize. Then, any time we want to find the sum of two numbers, we don't have to manipulate sets of objects again. We merely consult the table.

Using the addition table instead of adding on our fingers is more than just a time-saving convenience. It is an act of abstraction that *has changed the meaning of addition.* When we add on our fingers, we are actually working with cardinal numbers, which are properties of sets of objects. When we use the addition table, we are performing an operation on abstract symbols. This operation can now be performed without regard to what the symbols stood for in the first place. Let us examine this abstract operation a little more closely.

We have a set of symbols, 1, 2, 3, 4, and so on, that we call numbers. If we pick any one of them, and then pick another one, the addition table assigns to this pair of numbers another number called its sum. For example, if we pick the number 2 first, and the number 3 next, the table assigns as their sum the number 5. We have picked the pair numbers 2 and 3 in a definite order, specifying that the and the 3 is second. So we can refer to it as the (2, 3). What the addition table does is assign mber called the sum to every ordered pair of s we describe it in this way, we recognize s *a mapping*. It maps the set of ordered pairs o the set of single numbers. We usually show g in a series of statements like this: $2+3=5$, 6, $2+5=7$, etc. Its nature as a mapping shows e clearly if we write it this way:

$$(2,3) \xrightarrow{+} 5$$

$$(2,4) \xrightarrow{+} 6$$

$$(2,5) \xrightarrow{+} 7$$

etc.

The plus sign printed over the arrows is a reminder that the mapping shown is the special mapping called addition. It is only one of a multitude of possible mappings. We could, if we wished, set up a mapping of pairs of numbers into the letters of the alphabet. If we wished, we could map pairs of numbers into people's names. We have a wide choice of mappings, because we can map any set of objects into any other set of objects in any way we please. The plus sign symbolizing addition calls our attention to the fact that it is a very special mapping with special characteristics. Now let us examine some of these characteristics.

We observe, first, that the numbers we are pairing off in our ordered pairs are selected from the list of symbols, 1, 2, 3, 4, etc. To distinguish this set of symbols from the cardinal numbers from which they were derived, let us give it a name. We shall call it the system of *natural numbers*. We observe next that the number assigned as sum to each ordered pair is selected *from the same set*, the system of natural numbers. A mapping which assigns to each ordered pair of objects in a set another object selected from the same set is called a *binary operation*. So addition is an [illegible] ample of a binary operation defined on the system [illegible] numbers.

If we consult the addition table, we find th[illegible] $2+3=5$, $3+2=5$, and $4+1=5$. The dif[illegible] pairs, (1, 4), (2, 3), (3, 2), and (4, 1) are al[illegible] the same image, 5. So addition is a many-to-o[illegible] In particular, an ordered pair like (2, 3) and the [illegible] obtained by having the 2 and 3 change places, bo[illegible] the same image. We could write $2+3=3+2$. A s[illegible] statement is true for the sum of every ordered pair of nat[illegible] numbers. We find that $5+2=2+5$, $9+16=16+$[illegible], etc. This characteristic of the addition of natural numbers can be summarized in the following rule: If the letter a stands for any natural number, and the letter b stands for any natural number, then $a+b=b+a$. That is, if the natural numbers being added *commute* or change places, the sum is still the same. So this rule is known as the *commutative law of addition*, and we say that addition of natural numbers is a commutative operation.

We are so accustomed to using the commutative law of addition that it may seem to be obvious, and hardly worth mentioning. But it needs special mention because, while some binary operations, like addition of natural numbers, obey a commutative law, there are others that do not. For example, one of the operations we learned in elementary school arithmetic is called *division*, and is denoted by the symbol ÷. It is not a commutative operation, because the numbers being divided cannot, in general, change places without changing the result. For example, $8 \div 2$ is not equal to $2 \div 8$.

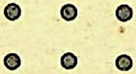
Addition, as we have talked about it so far, is an operation performed on a *pair* of numbers. We can also extend it to three numbers. We can add three numbers by first adding two of them, and then adding the sum to the third. However, for three numbers like 2, 3, and 7, listed in a definite order, we have a choice of two ways of doing it. We
ght add the sum of 2 and 3 to 7, or we might add 2 to
3 and 7. These two possibilities can be written
is form: $(2+3)+7$, and $2+(3+7)$. In this
parentheses indicate which sum is to be found
e carry out these additions, we find that it
any difference which sum is found first, be-
esults come out the same: $(2+3)+7=5+$
nd $2+(3+7)=2+10=12$. This is a charac-
of the addition of three natural numbers, no matter
numbers are used. It is expressed in the rule, $(a+b)$
$c=a+(b+c)$, where a, b, and c stand for any natural numbers. This rule says that in the first step of the addition we are free to associate the middle number either with the number on the left or with the number on the right. So the rule is known as the *associative law of addition,* and we say that addition of natural numbers is an associative operation. Since it makes no difference which pair of numbers we add first, we may as well leave out the parentheses altogether, and write the sum of a, b, and c as $a+b+c$, where it is understood that $a+b+c=a+(b+c)=(a+b)+c$.

The associative law, too, deserves special mention because it is a special property of addition of natural numbers, which it shares with some binary operations but not with all. For example, suppose we use the symbol *av* to designate the operation "take the average of." It is a binary operation that can be performed on the familiar whole numbers and fractions that we use every day. In this notation, 8 *av* 16 means the average of 8 and 16, which is 12. The symbol 12 *av* 12 means the average of 12 and 12, which is also 12. The symbol 16 *av* 12 means 14, and the symbol 8 *av* 14 means 11. This operation does not obey an associative law, because (8 *av* 16) *av* 12 is *not* equal to 8 *av* (16 *av* 12). In fact, (8 *av* 16) *av* 12 means 12 *av* 12, or 12, while 8 *av* (16 *av* 12) means 8 *av* 14, or 11.

By a step-by-step process, the use of the commutative law and the associative law for addition of natural numbers can be extended into a general rule for the sum of any finite selection of natural numbers: When you add a fini[illegible] selection of natural numbers, you can list the[illegible] order, and group them as you please. The sum w[illegible] come out the same.

Multiplication

The meaning of multiplication of natural nu[illegible] the meaning of addition, can be stated first i[illegible] sets of objects. To multiply 2 times 3, we set up a re[illegible] lar array of objects, consisting of two rows, with thr[illegible] jects in each row. Then we find the cardinal number [illegible]



this set. In general, to multiply the numbers *a* and *b*, we find the cardinal number of a set consisting of *a* rows with *b* objects in each row. The answer is called the product of *a* and *b*, and is designated by $a \cdot b$, where we use a dot as the symbol for multiplication. Once we have found the product of two natural numbers, we can record it for future reference in a multiplication table. Then we can separate the operation of multiplication from its original meaning of finding the cardinal number of a rectangular array of objects. We can think of it instead as merely a mapping of ordered pairs of natural numbers into the system of natural numbers. We usually show the mapping by a series of statements like this: $2 \cdot 3 = 6$, $2 \cdot 4 = 8$, $2 \cdot 5 = 10$, etc.

However, as in the case of addition, we can express it with the help of arrows:

$$(2, 3) \xrightarrow{\cdot} 6$$

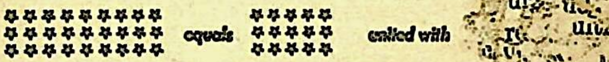
$$(2, 4) \xrightarrow{\cdot} 8$$

$$(2, 5) \xrightarrow{\cdot} 10$$

etc.

e the mapping is defined for every ordered pair of
mbers, and the image under the mapping is al-
tural number, multiplication, like addition, is a
tion on the system of natural numbers. We
ur experience with multiplication of natural
$2 \cdot 3 = 3 \cdot 2$, $2 \cdot 4 = 4 \cdot 2$, $2 \cdot 5 = 5 \cdot 2$, etc.
if a and b are any natural numbers, $a \cdot b =$
is known as the *commutative law of multiplica-*
ultiplication, like addition, also obeys an *associative*
$a \cdot (b \cdot c) = (a \cdot b) \cdot c$. This is seen, for example, in
e fact that $2 \cdot (3 \cdot 5) = 2 \cdot 15 = 30$, and $(2 \cdot 3) \cdot 5 =$
$6 \cdot 5 = 30$. Because of this law, we can write the product of three numbers without parentheses, and give it a definite meaning: $a \cdot b \cdot c = a \cdot (b \cdot c) = (a \cdot b) \cdot c$. Combining the commutative law and associative law of multiplication leads to the general rule: when you multiply a finite selection of natural numbers, you can list them in any order, and group them as you please. The product will always come out the same.

There is one more characteristic of the multiplication of natural numbers that links it with addition. We can understand it best by going back to the original meaning of multiplication as finding the cardinal number of a rectangular array, and the original meaning of addition as finding the cardinal number of a united set. Printed below is a rectangular array of stars, consisting of three rows with nine stars in each row. The number of stars in this rectangular array is $3 \cdot 9$. Since a row of nine stars can be thought of as the first five stars united with four other stars, we can write 9 = 5 + 4. So the number of stars in the rectangular array can also be written as 3 · (5 + 4). Now, suppose we move the first five stars in each row over to the left, so that a space separates them from the rest of the stars in the same row. The effect is to split our rectangle into two rectangles. One rectangle has three rows with five stars in each row, so it contains 3 · 5 stars. The other rectangle has three rows with four stars in each row, so it contains 3 · 4 stars. Since we get the original rectangle by uniting the two smaller rectangles, the number of stars in the original rectangle is the sum of the numbers of stars in t[illegible] smaller rectangles. This fact is expressed in the s[illegible] that 3 · (5 + 4) = (3 · 5) + (3 · 4). We can ver[illegible]



rectness of the statement by noting that 3 · 9 [illegible] 15 + 12 = 27. In general, if a, b, and c, stand for [illegible] numbers, $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$. Similarly, [illegible] $c) \cdot a = (b \cdot a) + (c \cdot a)$. This rule is known as the *distributive law* and expresses the fact that multiplication is distributive with respect to addition. That is, the multiplier can be distributed among the individual terms in the expression it multiplies. In the statement of this law, multiplication and addition cannot change places. While 3 + (5 · 4) = 3 + 20 = 23, (3 + 5) · (3 + 4) = 8 · 7 = 56, so that addition is *not* distributive with respect to multiplication. It is customary, in writing an expression like $(a \cdot b) + (a \cdot c)$ to leave the parentheses out, so that it looks like this: $a \cdot b + a \cdot c$. In such an expression, which gives instructions for doing both multiplication and addition of some numbers, it is understood that the multiplications must be done first.

The Five Laws

We originally introduced the natural numbers as symbols for the cardinal numbers. Then we made these observations about them: There are two binary operations defined on the natural number system, and we call them addition and multiplication. The properties of these operations are embodied in the addition and multiplication tables. By examining these tables, we found five laws that are obeyed by the natural number system: the commutative and associative laws of addition, the commutative and associative laws of multiplication, and the distributive law which asserts that multiplication is distributive with respect t addition. These laws have a special significance in the de velopment of our notion of what a number is. We find tha when we carry out computations with numbers we do no to keep in mind their original meaning as cardina It is enough to think of them as abstract symbol each other by addition and multiplication table these five laws. This fact suggests that we re ers as follows: *A number system is any collec jects on which two binary operations calle and multiplication are defined, such that additio mutative and associative, multiplication is commuta and associative, and multiplication is distributive wit respect to addition.*

This definition is a declaration of independence for th idea of a number system. It frees it from its cardinal number ancestry and permits it to lead its own life. I allows it to expand and grow. When a number system i defined in this way, we find that there is not just one num ber system, but many number systems. We find, too, tha it is possible for one number system to be part of a large number system, which in turn is part of a still larger nu ber system, and so on. In fact, the core of this book is th systematic construction of larger and larger number sy tems, using the natural numbers as a foundation. At eac stage of the construction we shall recognize that we hav built a number system when we find that it has two binar operations that obey the five laws:

I. $a + b = b + a$

II. $(a + b) + c = a + (b + c)$

III. $a \cdot b = b \cdot a$

$$\text{IV.}\quad (a \cdot b) \cdot c = a \cdot (b \cdot c)$$

$$\text{V.}\quad a \cdot (b + c) = a \cdot b + a \cdot c$$

$$\text{or}\quad (b + c) \cdot a = b \cdot a + c \cdot a$$

Large and Small Numbers

The natural number system has some other important characteristics besides the five laws. One of these is that we can compare any two numbers in it for size. The number 5 is larger than 4, and 4 in turn is larger than 3. The notion of larger and smaller is derived from addition in this way. We say that b is larger than a if b is equal to the sum [illegible] and some other natural number. For example, 5 [illegible] than 4, because $5 = 4 + 1$; 5 is larger than [illegible] $= 3 + 2$.

One System with Many Disguises

There are many different ways of writing the [illegible] numbers. In the system of Arabic numerals that w[illegible] every day, the numbers one, two, three, four and five [illegible] written as 1, 2, 3, 4, 5. In Roman numerals, still used on clock faces and monuments, they are written as I, II, III, IV, V. In Hebrew they are written as the first five letters of the alphabet. If we think of these different systems of numerals as symbols for the cardinal numbers, then they are different ways of representing one and the same number system. However, we may also think of each system of numerals as a separate number system in its own right, with addition and multiplication defined by its addition and multiplication tables. The Arabic, the Roman, and the Hebrew numerals could then be referred to legitimately as three separate number systems. But they are number systems that can be used interchangeably, so, although they are separate systems, they are still somehow the same. In the next chapter we shall encounter number systems that are not interchangeable and may not be considered the same. In order to recognize when number systems are interchangeable, and when they are not, we have to define what we mean when we say different systems are the same.

What we have in mind is that they have the *same structu* For two number systems to have the same structure, ea number in one system must have a counterpart in th other system. We can express this requirement in technic language by saying that there must be a mapping of o system into the other that places them in one-to-one co respondence. But the one-to-one correspondence alone not enough. We want to be sure, too, that the results computations in one system will correspond to the resu f computations in the other system. So we say that t her systems have the same structure, or are *isomorph* ere is a mapping of one into the other that p one-to-one correspondence, and (2) under t ums and products are preserved. The requi also be stated in this way: Under the mappi ment in one system has an image in the oth ver, the image of the sum of two numbers is the s e images; and the image of the product is the prod of the images. Comparing Arabic numerals and Rom numerals, for example, we can set up a one-to-one co respondence, shown in part in this table:

$$\begin{array}{c} 1 \longleftrightarrow \mathrm{I} \\ 2 \longleftrightarrow \mathrm{II} \\ 3 \longleftrightarrow \mathrm{III} \\ 4 \longleftrightarrow \mathrm{IV} \\ 5 \longleftrightarrow \mathrm{V} \\ 6 \longleftrightarrow \mathrm{VI} \end{array}$$

Each system has its own addition and multiplication tab part of which is shown in the customary square arran ments below:

Addition

+	1	2	3
1	2	3	4
2	3	4	5
3	4	5	6

+	I	II	I
I	II	III	I
II	III	IV	
III	IV	V	V

ultiplication

•	1	2	3
1	1	2	3
2	2	4	6
3	3	6	9

•	I	II	III
I	I	II	III
II	II	IV	VI
III	III	VI	IX

nder the mapping the image of 2 is II, and the image of 3
III. The sum of 2 and 3 is 5. The sum of II and II[illegible]
which is the image of 5. So the sum of the image[illegible]
age of the sum. The product of 2 and 3 is 6. The [illegible]
II and III is VI, which is the image of 6. So [illegible]
t of the images is the image of the product. [illegible]
erals and Roman numerals, considered as separ[illegible]
r systems, are isomorphic to each other. A[illegible]
mbers in one system look different from numbers i[illegible]
her system, the relationships within the systems, as [illegible]
essed in the addition and multiplication tables, have the
me structure. So the two systems are really only one
ucture appearing in two different styles of dress.

ro and One

Arabic numerals displaced all others because of their
eat convenience. They are most convenient to use be-
use they give us a way of writing an indefinite amount of
mbers while using only a small number of symbols called
gits. This feat is accomplished by attaching different
eanings to the same digit. In the number 111, three one's
e used, and each has a different meaning. The 1 on the
treme right stands for the number *one*. The 1 in the
cond column from the right stands for the number *ten*,
d the 1 in the third column stands for the number *one*
ndred. The symbol stands for the sum of one, ten, and
e hundred. Because the meaning of a digit depends on
position in the written numeral, we say the Arabic sys-
m of numerals is a *place value* system. To represent
ree *hundreds* plus two *tens* plus five *ones*, we write 325.

Now suppose we want to write the symbol for three te We put a 3 into the second column from the left. But w won't recognize it as the second column unless we wri something down in the first column. This makes it nece sary to think of three tens as three tens plus *no ones*, an to introduce a symbol to represent the absence of ones. W use the symbol 0 for this purpose, and call it zero. Th concept of a number representing *none* was first conceiv by the Hindus, and was later taken over by the Arabs a built into their system of numerals. Zero became a ne ber in the natural number system, and had to be i ed into the addition and multiplication tables ch is consistent with the rest of the tables. Th by using these rules for computation w plus any number gives that number aga times any number gives zero. The first of th n be written in symbols as follows: $0 + x = x$, natural number x. In later chapters, we shall be bu up some other number systems. It will be necessary us to find out whether these number systems contain element that behaves like the zero of the natural num system. In our search for a zero element, we shall use t first rule as our criterion. If a number system contains number a such that $a + x = x$, for all numbers x in th system, then we shall call a a *zero element*.

The distinguishing feature of a zero element is th *adding* it to another number leaves that number unchang There is a natural number that has the same relationship multiplication that zero has to addition. This number the number *one*, which obeys the rule: $1 \cdot x = x$, for natural numbers x. That is, *multiplying* a number by leaves that number unchanged. In the number systems explore later, we shall sometimes find an element that this property, and shall call it a *unity element*.

When we write out sums, we always use the symbol to stand for the operation of addition. We could, if wished, use some other symbol instead, as long as we agr on its meaning. We might, for example, use the symbo

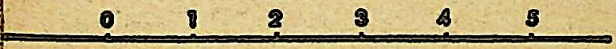
represent the operation. In that case, the characteristic
perty of 0 could be written in this way: $0 * x = x$. In
same way, we could, if we wished, change the symbol
multiplication. If, temporarily, we used the symbol *
represent multiplication, then the characteristic property
1 would be written as follows: $1 * x = x$.

The similarity in form of these two statements empha-
es the fact that 0 and 1 really both have the same prop-
y, except that each has it in relation to another opera-
n. They are both examples of what is known as an
ntity element. In any system in which a binary oper [illegible]
defined, and is symbolized by *, if there is an el [illegible]
at has the property $e * x = x$, for all values [illegible]
stem, then e is called an identity element. Th [illegible]
used in this definition of an identity element bec [illegible]
initial letter of the German word *einheit*, which [illegible]
ity.

Now we can state more precisely how the terms z [illegible]
ment and *unity element* are used in mathematics today.
henever a binary operation is denoted by the symbol +
d is called "plus," the identity element for that operation
called a *zero element*, and is denoted by 0. Whenever a
nary operation is denoted by the symbol ·, and is called
mes," the identity element for that operation is called a
ity element and is denoted by 1. We shall use this con-
ntion many times in later chapters.

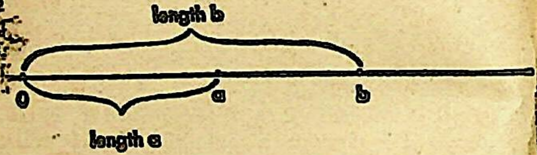
ints on a Line

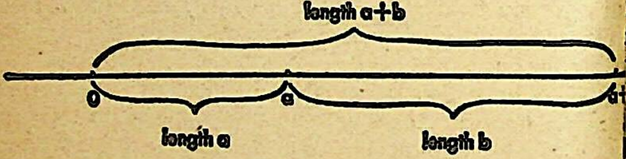
It is possible to represent the natural numbers as points
a line. On any straight line, choose a point and call it 0.
is point divides the line into two half-lines. On one of
se half-lines, choose another point and call it 1. Now
ntinue locating points further and further away from 0
making the distance from each point to the next one
same length as the distance from 0 to 1. Label these



new points successively 2, 3, 4, 5, etc. We then have endless sequence of points that is in one-to-one cor spondence with the natural number system. The numb attached to each point is its distance from 0, expressed terms of the distance from 0 to 1 as the unit of length.

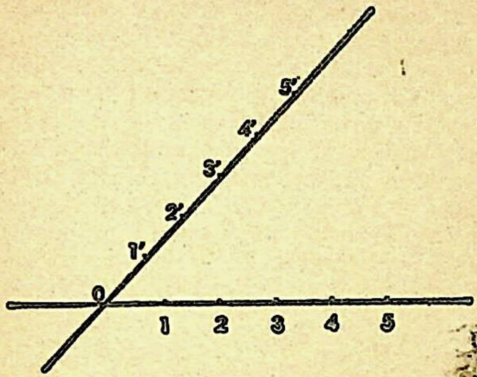
We can define addition and multiplication for the points by means of geometric constructions. Here, for e ample, is one way of doing it: To add a and b, measure o from a, in the direction away from 0, a length equal the distance from 0 to b. The point located in this w distance from 0 equal to $a + b$. To multiply a and





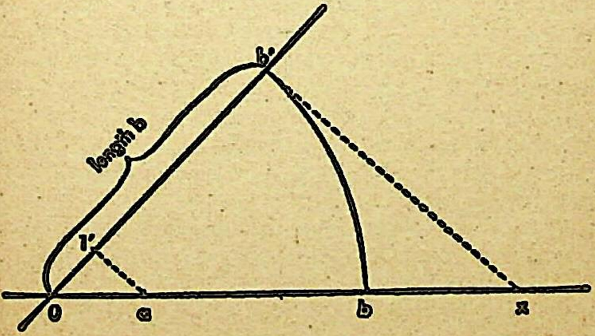
first draw another line intersecting this one at 0. Use same scheme for assigning numbers to points on this li Locate points 1′, 2′, 3′, etc., on the line, so that success points are separated by equal distances, all equal to distance from 0 to 1′. Join 1′ on the new line to a on original line. Locate b' on the new line at a distance fr 0 equal to b. Then through b' draw a line parallel to line just drawn from 1′ to a. It will cross the original l at a point that will represent $a \cdot b$.

The construction for addition obviously corresponds ordinary addition of numbers. The construction for mu plication corresponds to ordinary multiplication for reason: If we designate by x the point that we have defi



s the product of a and b, then x is its distance fr [illegible] riangles $(0\ 1'\ a)$ and $(0\ b'\ x)$ are similar, so thei [illegible] ponding sides are proportional. Then $1 : b = a : x$. [illegible] his proportion, we find that $x = a \cdot b$. With addition an ultiplication defined by these constructions, the system f points on the half-line is isomorphic to the natural umber system.

After we have assigned numbers in this way to points on line, we find that there are still many points on the line hat do not have numbers. All of our numbers are on one ide of 0. There are none on the other side. Moreover, we



have not assigned numbers to the points between th that represent successive natural numbers. For examp there are no numbers assigned to the points that lie tween 0 and 1. This is a defect that will be remedied s by step as we go along. One of our chief purposes will to build up a number system that has enough points it for us to assign a number for every point on the line. hope to do it, too, in such a way that the expanded num system and the whole line will be isomorphic.

The Natural Numbers

far we have encountered several different syst capable of representing the natural numbers. consists of the common Arabic numerals, ition and multiplication tables. A second of the Roman numerals, with their tables. system consists of points on a line, and the appro constructions for adding or multiplying them. variety of representations raises the question, "What is natural number system, anyhow?"

We might try to answer this question by listing so characteristics that all these number systems have in co mon. All three, for example, obey the five laws. But is not an adequate answer, because *all* number syst as we have defined the term, will obey these five laws. we intend to produce some number systems that are interchangeable with the natural number system at all. define the natural number system, we must list not me characteristics that all of its representations have in c mon. We must note particularly its *distinguishing* c acteristics. This is done by choosing the defining charac istics in such a way that all systems that have th characteristics must be isomorphic to each other. Suc selection of characteristics that effectively defines one only one structure is called a system of axioms for structure. Here is a system of axioms for the natural nu ber system (not including 0), first formulated by the mat matician Peano:

A set of elements is called a natural number system if it has the following characteristics:

(1) It contains an element called 1.

(2) For every member in the system, there is another member (and only one) called its successor.

(3) Two distinct members do not have the same successor.

(4) There is no member of the system that has 1 as its successor.

(5) If a set of elements belonging to the system contains 1, and, for each member that it contains, also [illegible]tains its successor, then this set contains t[illegible] system.

Notice that addition and multiplication ar[illegible]tioned in these axioms at all. Peano defined these [illegible] in terms of his axioms as follows: For any natural [illegible] x and y,

let $x+1=$ the successor of x;
let $x+$ (the successor of y) $=$ the successor of $(x+y)$;
let $x \cdot 1=x$;
let $x \cdot$ (the successor of y) $=x \cdot y+x$.

With these definitions it is possible to prove that the natural number system obeys the five laws.

What Peano did for the natural number system is typical of the way in which mathematical structures are studied today. In modern mathematics, a mathematical structure is often defined as a set of objects that satisfies a specified set of axioms. If the structure defined is to be unique, the axioms are chosen so that all systems that satisfy the axioms will be isomorphic to each other. Different sets of axioms have been formulated for the various mathematical structures needed in practical applications.

DO IT YOURSELF

. By using double-headed arrows, as on page 26, set up a one-to-one correspondence between the numbers 1, 2, 3, 4, 5 and the letters a, e, i, o, u.

2. An addition operation for the system consisting of tw elements, a and b, is defined by the following table:

$+$	a	b
a	a	b
b	b	a

a) Does this system have a zero element?

b) Show that addition is commutative in this system.

Verify from the table that $a + (a + b) = (a +$
b.

symbol M stand for the binary operation, "ta
imum of." For example, $5 M 7$ means 7; $8 M$
8; $6 M 6$ means 6. Compare $8 M 3$ with $3 M$
and b are any two natural numbers, compare $a M$
$b M a$. Is the operation M commutative? Comp
$8 M (3 M 7)$ with $(8 M 3) M 7$. In general, if a, b, an
are any three natural numbers, compare $a M (b M$
with $(a M b) M c$. Is the operation M associative?

4. In Ex. 2, an addition operation is defined for the whose elements are a and b. If this operation obeys associative law, then $x + (y + z) = (x + y) + z$ for values of x, y and z.
   a) The associative law stated above is an abbreviat way of making eight separate statements, obtain by replacing x, y and z by a or b. Write these eig statements.
   b) Prove the associative law for this system by verifyi from the table that all eight statements are true.

CHAPTER II

Number Systems without "Numbers"

THE word "number," as we use it in eve [illegible] efers to a symbol associated with counting or [illegible] Ve have broken away from this usage in the de [illegible] umber system given in Chapter I. We defined [illegible] ystem as any set of objects on which two binary op [illegible] re defined that obey the five laws listed on pages 24- [illegible] his definition, there is no reference to counting or measur ng. The five laws are concerned only with the way in which he numbers are related to each other by the addition and nultiplication tables. To emphasize this fact, we separated he concept of natural number from that of cardinal num er. While cardinal numbers are properties of actual sets f objects, and are intrinsically related to counting, natural umbers are abstract symbols whose entire meaning lies in he formal rules by which we manipulate them.

Nevertheless, the natural numbers do not convey the full neaning of our break with common usage in the definition f number. Introducing the natural numbers effected a eparation from the cardinal numbers, but not a divorce. he cardinal number system is still lurking in the back round, because it is isomorphic to the natural number ystem. This fact may arouse the suspicion that no signifi ant change in the concept of number has really been intro uced, and that numbers are still essentially bound up ith counting and measuring. However, a real change *has* een introduced by our definition of number system. The

purpose of this chapter is to demonstrate this fact convi
ingly by producing some number systems without "nu
bers." These number systems will consist of elements t
have no direct relationship to counting or measuring, a
so are not "numbers" in the sense in which the word
commonly used. However, they will be genuine numb
systems in the sense of our definition.

Subsets of a Set

We shall construct these number systems with the h
he simple notion of a set. A *set* is any collection
The objects that belong to a set are called
A set is defined by specifying which objects
f the set. This may be done by stating some
f the elements can be identified, or by actu
the elements on display. The symbol comm
or a set is a pair of braces, with the elements
the set on display inside, or with the rule by which they
identified printed inside. For example, here is a set defi
by a rule:

{natural numbers larger than 4, but less than 10}

The same set can be represented by putting its elements
display:

{5, 6, 7, 8, 9}

Other sets can be formed from a given set by remov
some of its elements. For example, if we remove the
ments 5, 7, and 8 from the set shown above, we are left w
the set {6, 9}. It is convenient to extend the notion of se
include what is left if we remove *all* the elements.
will be a "set" with no elements in it, and will be refe
to as the *empty set*. To symbolize it, we shall show a
of braces with no elements on display inside. A set obta
by removing none, or some, or all of the elements of a gi
set is called a *subset* of that set. For example, the
{x, y, z} has eight subsets, listed below:

$$\{x, y, z\} \quad \{x, y\} \quad \{x, z\} \quad \{y, z\}$$
$$\{x\} \quad \{y\} \quad \{z\} \quad \{\ \}$$

Notice that the given set is one of its own subsets, and the empty set is one of the subsets, too.

Operations on Subsets

To define a number system, we must first specify what the elements of the number system are. We shall use as elements all the subsets of a given set. As a specific example, let us build a number system out of the subsets of the [illegible] $\{x, y, z\}$. For convenience in talking about the [illegible] assign a name to each of these subsets. We shall [illegible] letters for their names, as follows:

$$\begin{array}{ll} I = \{x, y, z\} & D = \{x\} \\ A = \{x, y\} & E = \{y\} \\ B = \{x, z\} & F = \{z\} \\ C = \{y, z\} & 0 = \{\ \} \end{array}$$

The symbols I and 0 are included among the names used for reasons that will become clear later.

The next step is to define two binary operations on these elements. A binary operation is defined when we set up some rule for assigning to each ordered pair of subsets some particular subset in the same list. We define the operation of forming a *union* of two subsets by means of this rule: The union of two subsets is another subset formed by taking as its elements those elements that are in *one or the other* of the subsets being united. For example, A contains the elements x and y. B contains the elements x and z. The elements that are in one or the other are x, y, and z. So the union of A and B is the set $\{x, y, z\}$, which we have called I. The union operation will be the "addition" operation of this number system. However, we shall not use the plus sign to represent it. Instead, we shall use the symbol $\cup$. The union of A and B will be written as $A \cup B$, and is read as "A union B." We have seen that $A \cup B = I$. The method of finding the union of two subsets will be clear from th following examples:

$$I \cup C = \{x, y, z\} \cup \{y, z\} = \{x, y, z\} = I$$

$$D \cup E = \{x\} \cup \{y\} = \{x, y\} = A$$

$$C \cup 0 = \{y, z\} \cup \{\ \} = \{y, z\} = C$$

The results of forming all possible unions can be summarized in this table of unions (the addition table for th number system we are constructing):

∪	*I*	*A*	*B*	*C*	*D*	*E*	*F*	0
[illegible]	*I*	*I*	*I*	*I*	*I*	*I*	*I*	*I*
[illegible]	*I*	*A*	*I*	*I*	*A*	*A*	*I*	*A*
B	*I*	*I*	*B*	*I*	*B*	*I*	*B*	*B*
C	*I*	*I*	*I*	*C*	*I*	*C*	*C*	*C*
D	*I*	*A*	*B*	*I*	*D*	*A*	*B*	*D*
E	*I*	*A*	*I*	*C*	*A*	*E*	*C*	*E*
F	*I*	*I*	*B*	*C*	*B*	*C*	*F*	*F*
0	*I*	*A*	*B*	*C*	*D*	*E*	*F*	0

The second binary operation we define is that of forming the *intersection* of two subsets. The intersection two subsets is another subset formed by taking as its el ments all those elements that are *in both* of the subsets bei intersected. For example, A contains the elements x and B contains the elements x and z. Only the element x is both A and B. So the intersection of A and B is the subs $\{x\}$, which we have called D. The intersection operatio will be the "multiplication" operation of the number syste we are constructing. We shall designate it by the symbol

The intersection of A and B will be written as $A \cap B$, and is read as "A intersection B." Then we see that $A \cap B = D$. When two subsets have no elements in common, their intersection is the empty set. The method of finding intersections is shown in the following examples:

$$\begin{aligned} I \cap C &= \{x, y, z\} \cap \{y, z\} = \{y, z\} = C \\ A \cap D &= \{x, y\} \cap \{x\} = \{x\} = D \\ B \cap 0 &= \{x, z\} \cap \{\ \} = \{\ \} = 0 \\ E \cap F &= \{y\} \cap \{z\} = \{\ \} = 0 \end{aligned}$$

The results of forming all possible intersections can b[illegible] marized in this table of intersections (the mul[illegible]g table for the number system we are constructi[illegible]

$\cap$	I	A	B	C	D	E	F	[illegible]
I	I	A	B	C	D	E	F	0
A	A	A	D	E	D	E	0	0
B	B	D	B	F	D	0	F	0
C	C	E	F	C	0	E	F	0
D	D	D	D	0	D	0	0	0
E	E	E	0	E	0	E	0	0
F	F	0	F	F	0	0	F	0
0	0	0	0	0	0	0	0	0

The Five Laws Are Obeyed

The operations "union" and "intersection" obey the five laws listed on pages 24-5. We can verify this fact by referring back to the meaning of these operations. Let us examine the laws one at a time, to see if they are obeyed.

1. *The commutative law of addition.* Union is our additive operation, so we must see whether $X \cup Y = Y \cup X$, wh X and Y represent any subsets of I. $X \cup Y$ means the that consists of those elements that are in X or Y. $Y \cup$ means the set that consists of those elements that are Y or X. These are obviously the same sets, so law numbe is obeyed.

2. *The associative law for unions.* We must see whethe $(X \cup Y) \cup Z = X \cup (Y \cup Z)$. The set $(X \cup Y) \cup Z$ the set consisting of elements that are in X or Y, or in The set $X \cup (Y \cup Z)$ is the set consisting of elements t in X, or in Y or Z. These are obviously the same s aw number 2 is obeyed.

commutative law of multiplication. Intersection plication operation, so we must see wheth $Y \cap X$. $X \cap Y$ means the set consisting of e at are in both X and Y. $Y \cap X$ means the set co ng of elements that are in both Y and X. These a bviously the same sets, so law number 3 is obeyed.

4. *The associative law for intersections.* $(X \cap Y) \cap$ means the set consisting of elements that are in X and and also in Z. $X \cap (Y \cap Z)$ means the set consisting elements that are in X, and also in Y and Z. Clearly, th $(X \cap Y) \cap Z = X \cap (Y \cap Z)$, and law number 4 obeyed.

5. *The distributive law.* We must see if $X \cap (Y \cup Z)$ $(X \cap Y) \cup (X \cap Z)$. $X \cap (Y \cup Z)$ means the set consis ing of elements that are in X and in Y or Z. $(X \cap Y)$ $(X \cap Z)$ means the set consisting of elements that are X and Y, or in X and Z. These are clearly the same se so law number 5 is obeyed.

Since the five laws are obeyed, the system of subsets I, with the operations union and intersection, forms a nu ber system. Similar number systems can be construct from the subsets of any given set. In the example just give we started with a set that contains three elements, a found that it has eight subsets. As a result we obtained number system that contained exactly eight members. H we started with a different number of elements, we would have obtained a number system with a different number of members. For example, a set with two elements has four subsets. A set with four elements has sixteen subsets. A set with five elements has thirty-two subsets. In general, a set with n elements has 2^n subsets.

Zero and Unity Elements

The number system we have constructed has a zero element and a unity element. Since union is our addition operation, a zero element would have to have the property th[illegible] when it is united with any element of the system, [illegible] that element unchanged. A glance at the union [illegible] page 38 shows that the empty set has this prop[illegible] is why we used the symbol 0 to represent it. Sin[illegible] tion is our multiplication operation, a unity eleme[illegible] have to have the property that when it is intersect[illegible] any element of the system, it leaves that element [illegible] changed. A glance at the intersection table on page 39 shows that the original set I has this property. We chose the symbol I to represent it because of its resemblance to the number 1.

Special Properties

The number system we have just constructed out of the subsets of $\{x, y, z\}$ has, as we have seen, some properties that it shares with the natural number system. These include obedience to the five laws, and possession of a zero element and a unity element. However, it also has some peculiar properties that are entirely unlike the properties of the natural number system. A few of these are noted here.

1. We can see from the tables that for any element X in the system, $X \cup X = X$, and $X \cap X = X$. That is, a subset united with itself yields the same subset, and a subset intersected with itself yields the same subset. In the natural number system, such an outcome is the exception rather than

the rule. $0 + 0 = 0$, but $2 + 2$ is not 2. $1 \cdot 1 = 1$, but 2
is not 2.

2. We have already observed that intersection is distri
tive with respect to union. It can also be verified that uni
is distributive with respect to intersection. That is, in
statement of the distributive law, union and intersect
can change places. This, too, is unlike what we found in
natural number system. There, while multiplication
distributive with respect to addition, addition is not d
tributive with respect to multiplication.

3. For each subset in the system, we can find anot
hat contains just those elements that the first one d
call this second subset the *complement* of the fi
use, while they do not overlap (their intersect
ether they complete the original set (their uni
X is any subset in the system, we denote its co
t by X'. In the system of subsets of $\{x, y, z\}$, A
$y\}$, so $A' = \{z\} = F$. Similarly, $B' = E$, $C' = D$,
$I' = 0$. The operation of "taking the complement" has
following properties:

$$X \cap X' = 0, X \cup X' = I; (X')' = X.$$

It also obeys the very useful law known as De Morga
Rule: The complement of a union is the intersection of
complements; and the complement of an intersection is
union of the complements. Written out in symbols, the
says:

$$(X \cup Y)' = X' \cap Y'; (X \cap Y)' = X' \cup Y'.$$

The truth of the law can be observed by noting that a un
consists of elements in one set *or* the other, an intersect
consists of elements in one set *and* another, and a comp
ment consists of the elements *not in* a particular set. T
De Morgan's rule says that "not in either X or Y" is
same as "not in X and not in Y"; and that "not in both
and Y" is the same as "not in X or not in Y." A little thou
will show that these statements are correct.

he Algebra of Logic

The number system we have constructed in this chapter only one of a whole family of number systems that have milar properties. They are called Boolean algebras. The pe of structure that they represent is not just a mathe-atical curiosity. It has an important practical application the study of logic, and in the design of electronic com-uters. In logic we study relationships among statements. he analysis of these relationships can be carried out in mbols in the following way: Let each proposition or state-ent be represented by a letter, such as p, q, or r. Use the mbol $\cup$ for "or," the symbol $\cap$ for "and," and the [illegible] for "not," as we already have done in the last paragraph. se 0 for any statement that is false, and I for any statement at is true. With this notation, the class of statements and eir logical relations becomes a Boolean algebra. Boolean gebras are named after the English mathematician, George oole, who pioneered in the study of symbolic logic.

DO IT YOURSELF

Assign names to the subsets of the set $\{x, y\}$ as follows: $I = \{x, y\}$, $A = \{x\}$, $B = \{y\}$, $0 = \{\ \}$ = the empty set.

a) Construct a table for the union operation for this system of subsets.

b) Construct a table for the intersection operation.

Let I represent the set $\{a, b, c, d, e, f, g, h\}$.
Let X represent the subset $\{a, b, c, d\}$.
Let Y represent the subset $\{a, b, e, f, g\}$.

a) What elements are in X', the complement of X in I?

b) What elements are in Y', the complement of Y in I?

c) What elements are in $X' \cup Y'$?

d) What elements are in $X \cap Y$?

e) What elements are in $(X \cap Y)'$?

f) Compare your answers to c) and e) to show that $X' \cup Y' = (X \cap Y)'$.

List all the sixteen subsets of the set $\{a, b, c, d\}$.

CHAPTER

New Numbers from Old

Questions That Have No Answers

IN OUR everyday use of the natural numbers, dding them and multiplying them, we someti sion to subtract them. The operation of subt be defined in terms of addition. The symbol 5 ks us the question, "What natural number adde 5?" Since the answer to the question is the nu , we say $5 - 3 = 2$. We call the answer the *difference* tween 5 and 3. The question can also be written in the f of an equation, $x + 3 = 5$, and the answer to the ques is the solution to the equation.

Our success in finding the difference between 5 a tempts us to try to find the difference between any natural numbers chosen at random. But then we run trouble. Suppose, for example, we try to find the differe between 3 and 5, written as $3 - 5$. First we have to inter the symbol as a question. It asks us, "What natural num added to 5 gives 3?" Unfortunately, the answer is that t isn't any such number. In the natural number system, cannot subtract any number from any other number. only time subtraction is possible in that system is when subtrahend is not larger than the minuend. If a and b st for any natural numbers, then the expression $a - b$ do always have a meaning. If we think of it as the quest "What natural number added to b gives a?", then it do always have an answer. If we use the equation $x + b$ which asks the same question, then it does not always h a solution. This is a defect of the natural number sys hat limits its usefulness. Because, although the question -5 is meaningless for the natural number system, there re practical problems that lead to just such a question. or example, if the temperature is 3 degrees, what will it e after the mercury drops 5 degrees? It would be useful o have a number system which contains a number that an serve as the answer to this question. The defect of the atural number system that we have observed confronts s with a challenge. Can we construct a number system hat does not have this defect? Can we build a number ystem in which subtraction is always possible for any pa[illegible] f numbers taken in any order, so that a $a-b$ alwa[illegible] meaning, and $x+b=a$ always has a solution? [illegible] hat we can.

Readers who have had high school algebra wil[illegible] hat a system of numbers that includes "negative" [illegible] s "positive" numbers is supposed to serve this p[illegible] But in their course in high school algebra, they were given his system as a finished product obeying certain mysterious ules such as, "the product of two negative numbers is a ositive number." In what follows, we do not take the xistence of such a number system for granted. We prove t exists by actually constructing it. We also remove the mystery surrounding its rules by actually deriving them rom the familiar rules governing the system of natural umbers.

Families of Differences

To construct the improved number system, we use a ather interesting device. The symbol $a-b$ asks us a question which does not always have an answer. To make sure that it will have an answer in the new system, we let the question be its own answer! In effect we say, let each expression like $5-3$, or $3-5$, or $2-7$, represent a number in the new system. To justify calling these strange things numbers we shall have to define addition and multiplication operations for them, and then show that with these operations they really constitute a number system.

However, we run into some complications even before w take our first step in this direction. In the natural numb system, $5 - 3$ does have an answer, and the answer is But $2 - 0$, $3 - 1$, $4 - 2$, $6 - 4$, and an endless list of simil symbols also represent 2. So we cannot simply let each su symbol stand for a separate number in the new system. W would want all of these symbols to represent the *same* nu ber, just as they do in the natural number system. We ta care of this difficulty by using as the elements of our n number system, not single symbols written in the for of a "difference" between two natural numbers, but who [illegible] of such differences. The first step is to establi [illegible] which we can recognize when two such symb [illegible] the same family. We get a clue to the rule [illegible] by examining the difference symbols that rep [illegible] number 2. The difference $3 - 1$ and the differe [illegible] represent the same number. Notice that if we a the left number of each symbol to the right number the other, we get the same sum: $3 + 4 = 6 + 1$. We sh use this relationship as the criterion for identifying diff ences that belong to the same family.

Now we are ready to carry out our construction step step. First we take all possible ordered pairs of natu numbers, such as 7 and 5, 3 and 9, 15 and 1, and so on. Th we write the "difference" of the numbers in the pair, tak in a definite order. Since the difference does not alw have a meaning in the natural number system, we shall use an ordinary minus sign when we write it. We shall the symbol ~ instead, to remind us that this is really subtraction of natural numbers, but merely a symbol s gested by subtraction. So we now have symbols like $7 \sim$ $3 \sim 9$, $15 \sim 1$, and so on, each of which will be calle "difference" between natural numbers.

Now we associate with each difference a whole fam of differences in the following way: *The family belong to a difference* $a \sim b$ *consists of all those differences* u *for which* $a + v = u + b$. To designate the family that longs to a difference we shall write that difference ins parentheses. Thus $(a \sim b)$ means the family of differences that belongs to $a \sim b$. The symbol $(3 \sim 1)$ means the family of differences that belongs to $3 \sim 1$. We have already seen that the difference $6 \sim 4$ belongs to this family because $+ 4 = 6 + 1$. We call these families of differences *integers*. They will be the elements of our new number system.

We observe immediately two characteristics of these families which we call integers:

1) A difference belongs to its own family. For example, $3 \sim 1$ belongs to $(3 \sim 1)$. This follows from the fact that $u \sim v$ belongs to $(a \sim b)$ if $a + v = u + b$. In this case $a = 3$, $b = 1$, $u = 3$, and $v = 1$, and $3 + 1 = 3 + 1$. In general, $a \sim b$ belongs to $(a \sim b)$ because $a + b = a + b$.

2) If one of two differences belongs to the family of the other, then they have the same families. Suppose, for example, that $a \sim b$ belongs to the family $(c \sim d)$. Then we can show that every member of $(a \sim b)$ belongs to $(c \sim d)$, and vice versa. If $p \sim q$ belongs to $(a \sim b)$, then by the criterion for membership in a family, $a + q = p + b$. However, $a \sim b$ belongs to $(c \sim d)$, so $c + b = a + d$. Adding these two equalities, we get $a + b + c + q = a + b + p + d$. Taking away $a + b$ from both sides, we get $c + q = p + d$. But this is equivalent to saying that $p \sim q$ belongs to $(c \sim d)$, according to our criterion for membership in a family. This shows that any member of $(a \sim b)$ also belongs to $(c \sim d)$. A similar argument, going through the same chain of steps in reverse, shows that every member of $(c \sim d)$ also belongs to $(a \sim b)$. So the families $(a \sim b)$ and $(c \sim d)$ have the same memberships, and are therefore the same.

The second characteristic of these families called integers has these consequences: First, each difference $a \sim b$ belongs to one and only one integer. Secondly, an integer may be represented by putting on display inside parentheses any one of the differences that belong to it. So $(3 \sim 1)$, $(4 \sim 2)$, $(5 \sim 3)$ all represent the same integer. Thirdly, the criterion for membership in an integer can also serve as a test for equality of integers. That is, the integers $(a \sim b)$ a $(c \sim d)$ are equal if and only if $a + d = c + b$. For exa ple, to prove that $(3 \sim 1) = (4 \sim 2)$, it is enough to serve that $3 + 2 = 4 + 1$.

Addition and Multiplication of Integers

Now we define addition and multiplication for the syst of integers. We assign a sum to any ordered pair of integ by means of the following defining equation:

$$(a \sim b) + (c \sim d) = (a + c \sim b + d)$$

The symbol on the right hand side of this equation rep an integer, because, if a and c are natural numb also a natural number. Similarly, $b + d$ is a natu Then $a + c \sim b + d$ is a difference of natural nu nd there is an integer that belongs to it.

We assign a product to any ordered pair of integers means of the following defining equation:

$$(a \sim b) \cdot (c \sim d) = (a \cdot c + b \cdot d \sim a \cdot d + b \cdot c)$$

Here, too, the right hand side represents an integer, beca the symbol inside the parentheses represents a differe of two natural numbers.

Notice that to find the sum of two integers, we make of the natural numbers whose differences are on disp inside the parentheses representing these integers. This leads to a problem that we have to pay attention to. E integer is a family of differences. Any member of a fam might be put on display to represent it. If we pick anot difference to represent each of the integers we are add will we still get the same sum? If we do not, our defini is useless. However, the definition was well chosen. If follow the directions it gives for finding the sum of integers, we arrive at *the same result* no matter which me bers of the integers are used to represent them. We s not take the time to prove this fact here, but shall me verify it for a particular case. Suppose we want to $(5 \sim 3)$ and $(6 \sim 5)$. Using our definition, we find that

$$(5 \sim 3) + (6 \sim 5) = (5 + 6 \sim 3 + 5) = (11 \sim 8).$$

owever, $(5 \sim 3)$ could also be represented by $(4 \sim 2)$, ecause $5 + 2 = 4 + 3$. Similarly, $(6 \sim 5)$ could also be presented by $(5 \sim 4)$, because $6 + 4 = 5 + 5$. If we apply r definition to these other representatives of the two tegers, we find that

$$(4 \sim 2) + (5 \sim 4) = (4 + 5 \sim 2 + 4) = (9 \sim 6).$$

y using different representatives for the two integers we ere adding, we got sums that look different. However, al- ough they look different, the sums are the same.

$$(11 \sim 8) = (9 \sim 6), \text{ because } 11 + 6 = 9 + 8.$$

The same problem arises in connection with our [illegible] on for multiplication of integers. The definition m[illegible] f a particular difference that belongs to each intege[illegible] can be shown that it does not matter which differe[illegible] hat belongs to an integer is chosen as its representative. hey all lead to the same product anyhow. So there is no mbiguity in our definitions of addition and multiplication.

he Integers Form a Number System

We now have a system of elements called integers, with n addition operation and a multiplication operation de- ned for this system. To show that the integers form a umber system, we have to prove that the operations obey he five laws listed on pages 24-5. As an example of how such proof is carried out, we give the details of the proof for he commutative law of addition. Let $(a \sim b)$ be any in- eger, and $(c \sim d)$ any other integer. We must show that $a \sim b) + (c \sim d) = (c \sim d) + (a \sim b)$. Applying our def- ition of addition of integers, we find that $(a \sim b) +$ $c \sim d) = (a + c \sim b + d)$, while $(c \sim d) + (a \sim b) =$ $c + a \sim d + b)$. But natural numbers obey the commuta- ve law for addition, so $a + c = c + a$, and $b + d = d + b$. his shows that $(a + c \sim b + d)$ and $(c + a \sim d + b)$ re the same integer. Therefore, $(a \sim b) + (c \sim d) =$ $c \sim d) + (a \sim b)$, and the commutative law for addition of integers is true. The other four laws are proved by sim arguments, using the definitions of addition and multipli tion of integers, and the fact that natural numbers known to obey the five laws.

Zero and Unity

Let us add the integer $(0 \sim 0)$ to any other inte $(a \sim b)$. Following the definition of addition, we find t $(0 \sim 0) + (a \sim b) = (0 + a \sim 0 + b) = (a \sim b)$, s $0 + a = a$, and $0 + b = b$. In other words, when $(0 \sim$ is added to any other integer, it leaves that integer nged. Therefore $(0 \sim 0)$ is a zero element for the sys gers. The integer $(0 \sim 0)$, like any other intege y of differences, and may be represented by any differences. We can identify what these differe like by using the criterion for belonging to an inte he difference $x \sim y$ belongs to the family of differe $(0 \sim 0)$, if and only if $0 + y = x + 0$, or $y = x$. Tha a difference belongs to the integer $(0 \sim 0)$ if and only i left number and the right number are equal. So $(1 \sim$ $(2 \sim 2)$, $(3 \sim 3)$, and so on, are other ways of writing zero element in the system of integers.

Let us see how the integer $(1 \sim 0)$ behaves under m plication. Following the definition of multiplication, find that $(1 \sim 0) \cdot (a \sim b) = (1 \cdot a + 0 \cdot b \sim 1 \cdot$ $0 \cdot a) = (a + 0 \sim b + 0) = (a \sim b)$. In other words, w $(1 \sim 0)$ is multiplied by any other integer, it leaves t integer unchanged. So $(1 \sim 0)$ is a unity element for system of integers. It may also be written in the f $(a + 1 \sim a)$, where a is any natural number. This fo from the test for equality of integers, because $a + 1 +$ $1 + a$.

The Negative of an Integer

Our purpose in constructing the system of integers to find a number system in which an equation of the f $X + B = A$ always has a solution. To show that we h achieved our purpose, we have to introduce a new conc concept of the *negative* of a number. We say that one mber is the negative of another if the sum of the two mbers is zero. In the natural number system, there is only e number that has a negative. That number is 0, and it its own negative, because $0 + 0 = 0$. No other natural mber has a negative, because if a natural number that is ferent from 0 is added to any other natural number, the m is different from 0. However, the system of integers is ogether different in this respect: In the system of integers, *every* number has a negative. We prove this fact by tually producing the negative of any integer. Let $(a \sim b)$ any integer. Then $(b \sim a)$ is its negative, because $\sim b) + (b \sim a) = (a + b \sim b + a) =$ the zero integer, cause, in the difference on display inside the parentheses, e left number is equal to the right number. Because [illegible] teger has a negative, we introduce a special symbol [illegible] ean "negative of." The minus sign is used for this purpose. A stands for an integer, then $-A$ is used to represent the gative of A.

Now we are prepared to show that the equation $+ B = A$ always has a solution, if X, A, and B stand for tegers. We have used capital letters here to represent tegers, so that we would not confuse them with natural mbers. Each integer is a family of differences of natural mbers, so the equation can also be written in this form: $\sim y) + (c \sim d) = (a \sim b)$. We solve the equation by ding the negative of $(c \sim d)$ to both sides. We get

$$(x \sim y) + (c \sim d) + (d \sim c) = (a \sim b) + (d \sim c).$$

ut the sum of $(c \sim d)$ and $(d \sim c)$ is the zero integer. And e zero integer added to $(x \sim y)$ leaves it unchanged. So e have $(x \sim y) = (a \sim b) + (d \sim c) = (a + d \sim b + c)$. r example, to solve $(x \sim y) + (8 \sim 1) = (3 \sim 2)$, add $\sim 8)$ to both sides. Then we get $(x \sim y) = (4 \sim 10)$.

On page 44 we saw that solving $X + B = A$ was another y of saying subtract B from A. Since, in the system of tegers, the equation always has a solution, it means that btraction is always possible in that system. In fact, our method for solving the equation suggests an approp definition for the subtraction of integers: To *subtrac* integer, means to add its negative. This definition m sense in the system of integers, where every number h negative. We could not have defined subtraction of nat numbers in the same way, because in the system of nat numbers, it is not true that every number has a nega

We Still Have the Natural Numbers

Among the integers, there are some special integers $(0 \sim 0)$, $(1 \sim 0)$, $(2 \sim 0)$, $(3 \sim 0)$, and so on, which ritten by putting on display a difference in which er on the right is 0. These special integers are ve integers. They can be placed in one-to-one ence with the natural numbers by pairing off 0 0), 1 with $(1 \sim 0)$, 2 with $(2 \sim 0)$, and so on. In al, in this correspondence, each natural number matched with the positive integer $(a \sim 0)$.

Now let us see what happens when we add or mul any two positive integers, $(a \sim 0)$ and $(b \sim 0)$. For sum, we get $(a \sim 0) + (b \sim 0) = (a + b \sim 0)$. So the of two positive integers is a positive integer. For product, we get $(a \sim 0) \cdot (b \sim 0) = (a \cdot b + 0 \cdot 0$ $0 + 0 \cdot b) = (a \cdot b \sim 0)$. So the product of two po integers is a positive integer. Moreover, the sum o integers matched with a and b is the integer matched $a + b$; and the product of the integers matched with b is the integer matched with $a \cdot b$. That is, under the to-one correspondence the sum of the images of two na numbers is the image of their sum; and the product o images is the image of the product.

So the positive integers are isomorphic to the na numbers. Because of the isomorphism, they can be use stead of the natural numbers, just as Roman num can be used instead of Arabic numerals. In this sense, we that the positive integers are the "same" as the na numbers. We take advantage of the isomorphism by u the notation for natural numbers as an abbreviated on for the positive integers. In this abbreviated notation represents $(0 \sim 0)$, 1 represents $(1 \sim 0)$, and so on.

Since the system of integers includes the positive integers hich are just "like" the natural numbers, they constitute extension of the natural number system. Using the system of integers instead of the system of natural numbers ves us a double advantage: We eliminate a defect of the tural number system, without losing the natural numbers emselves.

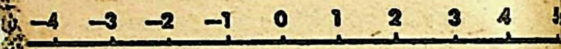
he Negative Integers

Every integer can be represented by a difference in which ther the left number or the right number is 0. This [illegible] one by simply subtracting from each of these [illegible] umbers the smaller of the two. The result will be [illegible] ice that belongs to the integer, and so can be used present it. For example, the integer $(8 \sim 3)$ is equal [illegible] $\sim 0)$. We know they are equal because $8 + 0 = 5 + 3$. he integer $(3 \sim 8)$ is equal to $(0 \sim 5)$, because $3 + 5 =$ $+ 8$. So every integer may be written in the form $(a \sim 0)$ $(0 \sim a)$. Those that can be written in the form $(a \sim 0)$ e the positive integers. Those that can be written in the rm $(0 \sim a)$ are called negative integers. Each of them is e negative of a positive integer. Using the minus sign as e symbol for "negative of," we get an abbreviated notan for them, too, by writing $-a$ for $(0 \sim a)$.

This is the form in which students are introduced to them the first time in courses in elementary algebra. The failiar rules for calculating with these symbols can all be rived from our definitions for the addition and multiication of integers. For example, the rule that the product two negative integers is a positive integer can be proved follows: $(0 \sim a) \cdot (0 \sim b) = (0 \cdot 0 + a \cdot b \sim 0 \cdot b +$ $0) = (a \cdot b \sim 0)$, which is positive.

The positive integers, being essentially carbon copies of natural numbers, can be represented as the natural mbers were on page 29, by points on the half-line that to the right of the 0. We can represent the negative

numbers pictorially, too, by putting them on the other
of the line, on the other side of 0. Arranging them on a
suggests that we can talk about larger and smaller inte
just as we were able to talk about larger and smaller nat
numbers. We give a meaning to the term larger as app
to integers by agreeing that one integer will be consid
larger than another if it lies to the right of it on the
on which the integers are represented as points in
diagram below. The term can also be defined without re
ence to the picture. If a and b are two different inte
we say that a is larger than b if $a - b$ is positive. It is un
stood here that $a - b$ means $a +$ the negative of b, in



rdance with the definition of subtraction of integers g
on page 52. On the basis of this definition, 2 is larger
-7, because $2 - (-7) = 2 + 7 = 9$, which is positive.
integer -4 is larger than -5, because $-4 - (-5) = -$
$5 = 1$, which is positive.

The Integers Form a Group

Before we proceed to any further extension of our n
ber system, let us stay with the system of integers f
while to observe some of its properties. We shall find w
the system of integers examples of some of the struct
that form the typical subject matter of modern ma
matics.

We have two binary operations defined for the syste
integers, addition and multiplication. Let us disregard m
tiplication, and list some of the properties the system ha
relation to the operation of addition alone. We obs
these characteristics: 1) The operation of addition is ass
ative. 2) The system contains a zero element. 3) For e
integer a in the system, its negative, $-a$, is also in
system. These characteristics make the system of inte
an example of the type of structure that is known

roup." In fact, with a slight change in notation, intro- ced in order to give them a more general form, these aracteristics make up the definition of a group.

As we did once before on page 28, let us designate the nary operation by the symbol *. Zero is an example of an entity element, and if we designate it now by the letter the property that $0+x=x+0=x$ can be written in e more general notation as follows: $e*x=x*e=x$. stead of the word "negative," let us substitute the word nverse," and designate the inverse of a by a^{-1}. In this tation, the fact that the sum of any integer and its nega- ve is zero takes this form: $a*a^{-1}=a^{-1}*a=e$.

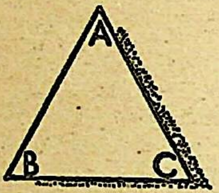
Here, then, is the definition of a group: A group is [illegible] m of elements for which a binary operation * is [illegible] d which has these properties: 1) The operation * [illegible] ative. 2) The system contains an identity element e, [illegible] e property that, if x is any element in the system, $e*x=$ $*e=x$. 3) For every element a in the system, its inverse $^{-1}$ is also in the system, with the property that $a*a^{-1}=$ $^{-1}*a=e$. The word "negative" is used for an inverse ly in the special case where the operation is designated + and is called "addition."

nother Example of a Group

The type of structure known as a group has been singled it for special study by mathematicians because it is found many places. The system of integers is only one of a ultitude of systems that have a group structure. It hap- ns to be a group that contains an infinite number of ements. But there are also groups that contain a finite mber of elements. As an example of a finite group, let examine the group of "symmetries" of an equilateral tri- gle.

An equilateral triangle has equal sides and equal angles. e symmetries of the triangle are motions that bring it to coincidence with itself. To get acquainted with these otions, it is best to see them in action as applied to a odel. So, before reading the paragraphs that follow, cut an equilateral triangle out of paper, and label its verti A, B, and C, as shown in the diagram. Enter the labels the reverse side of the paper, too, so that the vertices can identified even when the triangle is turned over.

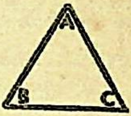
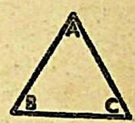
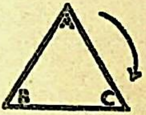
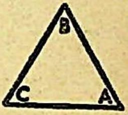
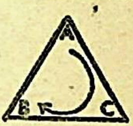
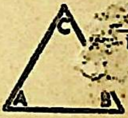
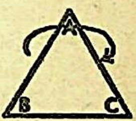
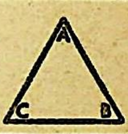
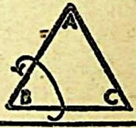
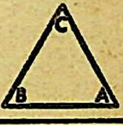
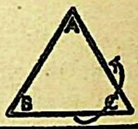
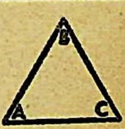
Let us begin with the triangle placed on a level surf so that one side, say BC, is horizontal, and the oppos vertex A lies above BC, as shown in the diagram belo



Now, as we discover motions that bring the triangle in coincidence with itself, we shall assign a name to each o One motion that will qualify is a clockwise rotation of 1 degrees around the center of the triangle. The result such a rotation is to put B in the place of A, A in the pla of C, and C in the place of B. Let us call this rotation Another motion that qualifies is a clockwise rotation of 2 degrees. It puts C in the place of A, B in the place of C, a A in the place of B. Let us call it Q. Another "motion" th qualifies is a clockwise rotation of 0 degrees. This, of cour involves no movement of the triangle at all, and leaves in the place of A, B in the place of B and C in the place C. We shall call it I.

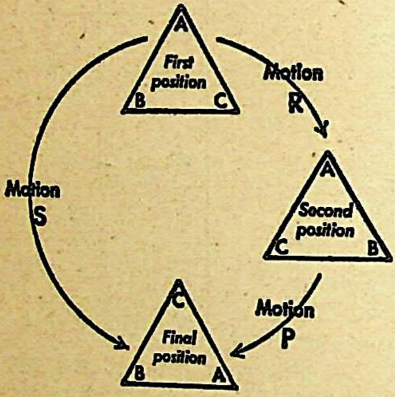
Before we go any further, let us make an agreement. W shall consider two motions as being the "same" if they ha the same effect. A clockwise rotation of 360 degrees has t same effect as a rotation of 0 degrees, so it, too, will called I. A counterclockwise rotation of 120 degrees has th same effect as a clockwise rotation of 240 degrees, so it, to will be called Q. Similarly, a counterclockwise rotation 240 degrees is the same as P.

There are three more motions that can bring the triang into coincidence with itself. In one of them, we flip the tr

Motion	Symbol	First position	Final position
No motion	I		
Rotate 120° clockwise	P		
Rotate 240° clockwise	Q		
Flip over, keeping top vertex fixed	R		
Flip over, keeping left vertex fixed	S		
Flip over, keeping right vertex fixed	T		

angle over to bring the bottom face up, while leaving the top vertex where it is. If we start with *A* at the top, *B* at the left, and *C* at the right, this motion puts *A* in its own place, and makes *B* and *C* change places. Let us call this motion *R*. A similar motion that keeps the vertex on the left fixed, while the top vertex and the vertex on the right ch places, will be called S. One which keeps the vertex on right fixed, while the top vertex and the vertex on the change places, will be called T.

We now have a system consisting of six elements, I, Q, R, S, and T. We define a binary operation $*$ for this tem as follows: If A and B represent any two of these tions, the product $A * B$ is the motion that results when two motions are performed one right after the other, wit



$P * R = S$

performed first, and A taking over where B leaves off. F example, $P * R$ means first perform R, then P. If we sta with vertex A at the top, B at the left, and C at the ri motion R leaves A at the top, and makes B and C ch places. Then, from this position, motion P rotates the angle 120 degrees clockwise. As a result, A moves to the sition on the right, B moves to the position on the left, C moves to the top. The effect is the same as if only single motion S had been used. So $P * R = S$. If we any two of the six motions at random, and perform one the other, we find that the result is always one of original six motions. The results of performing the op tion * can be summarized in the multiplication table shown below, where the motion that is performed first, and is written on the right hand side in a product, is listed at the top of the table, and the motion that is performed second, and is written on the left hand side in a product, is written at the side.

MULTIPLICATION TABLE
FOR SYMMETRIES OF THE TRIANGLE

*	*I*	*P*	*Q*	*R*	*S*	*T*
I	*I*	*P*	*Q*	*R*	*S*	*T*
P	*P*	*Q*	*I*	*S*	*T*	*R*
Q	*Q*	*I*	*P*	*T*	*R*	*S*
R	*R*	*T*	*S*	*I*	*Q*	*P*
S	*S*	*R*	*T*	*P*	*I*	*Q*
T	*T*	*S*	*R*	*Q*	*P*	*I*

To show that the symmetries of the triangle form a group, we have to prove that the three requirements for a group are satisfied. 1) The operation * is associative. That s, if A, B, and C are any three motions in the system, $A * (B * C) = (A * B) * C$. This can be verified from the table, or from the fact that both symbols, $A * (B * C)$, and $(A * B) * C$ require that C be performed first, B second, and A third, so that both lead to the same result. The parentheses here signify no more than a pause between motions. 2) The system has an identity element. In fact I is the identity element, because the table shows that if X stands for any element, $I * X = X * I = X$. 3) For every element in the system, there is an inverse element that is also in the system. The inverse of P is Q, and the inverse of Q is P, because $P * Q = Q * P = I$. Each of the elements, I, R, S, and T is its own inverse, because $I * I = I$, $R * R = I$,

$S * S = I$, and $T * T = I$. Therefore the symmetries of triangle, with the operation * as defined, form a group.

Not All Groups Are Commutative

We see from the multiplication table for the group symmetries that $P * R = S$, and that $R * P = T$. There $P * R$ is not equal to $R * P$. The operation * in this gr does not obey the commutative law. On the other hand the system of integers, which forms a group with the op tion +, the operation is commutative. A group wh obeys the commutative law is called a commutative gro It is also called an *abelian* group in honor of one of [illegible]eers in the study of groups, Niels Henrik Abel, who [illegible]e age of 27 in 1829. It is now the custom to use [illegible]bol + to represent the binary operation in a group if it is a commutative or abelian group. If the plus sig used to designate the operation, then the word *zero* is instead of *identity*, and the word *negative* is used instea *inverse*.

Although the system of integers is the most comm used system that has a group structure, it was not the system to have its group structure analyzed. The groups that were studied extensively were finite gro like the group of symmetries of the triangle. They came in the theory of equations, as studied by the methods of French mathematician Evariste Galois. Galois recorded findings in 1832 in a paper that he wrote the night be he was killed in a duel at the age of 21. After being lected for about thirty years, they were saved from scurity and extended by other mathematicians. Since the concept of a group has invaded every branch of ma matics.

The Integers Form a Ring

To observe the group structure of the system of inte we disregarded the operation of multiplication. Now, i put multiplication back into the picture, we find in system of integers an example of another structure plays an important part in modern mathematics, a *ring*. A system of elements is called a ring if there are two binary operations defined for the system, and they have these properties: 1) Both operations are associative. 2) The system is an abelian group with respect to one of the operations. This operation is designated by $+$, and is called addition. 3) The other operation is distributive with respect to addition. If we call it multiplication, and represent it by the usual symbol for "times," then the distributive law takes this form: $a \cdot (b+c) = a \cdot b + a \cdot c$; and $(b+c) \cdot a = b \cdot a + c \cdot a$. It is necessary to state the distributive law in two parts in this way, because the multiplication operation need not be commutative, so multiplication by a from the left is not the same as multiplication by a from the right.

The fact that the system of integers is a ring follows from the properties of the system that we have already seen.

Multiplication by Zero

One of the familiar rules of arithmetic that we use every day when we work with natural numbers is that zero times any number gives a product equal to zero. This rule turns out to be true in the system of integers as well. In fact, we can show that it has to be true *in any system that has a ring structure*. In any ring, if x is any element, and 0 is its zero element, $x + 0 = x$. Now multiply both sides of this equation by any other element in the ring, say y. Then we have $y \cdot (x+0) = y \cdot x$. By the distributive law in the ring, we can replace $y \cdot (x+0)$ by $y \cdot x + y \cdot 0$, so we get $y \cdot x + y \cdot 0 = y \cdot x$. Since the ring is a group with respect to addition, the element $y \cdot x$ has a negative, $-(y \cdot x)$. Let us add this negative to both sides of the equation. We get $-(y \cdot x) + y \cdot x + y \cdot 0 = -(y \cdot x) + y \cdot x$. But $-(y \cdot x) + y \cdot x = 0$, from the definition of a negative. So we now have $0 + y \cdot 0 = 0$. But since 0 is a zero element for addition, $0 + y \cdot 0$ can be replaced by $y \cdot 0$. This leads to the conclusion that $y \cdot 0 = 0$.

Hidden Groups and Rings

The group structure and ring structure that we have o served so far in the system of integers lie on the surfac we might say, because they embrace the system as a whol However, there are more group and ring structures hidd within the system of integers as substructures. We sha produce some examples for examination.

Suppose we divide the integer 6 by 3. The quotient i 2, which is also an integer. Then 6 can be written as 2 · There are other integers that can also be written as som integer times 3. For example, —9 = —3 · 3. All such integers are called integral multiples of 3. The integral multi [illegible] of 3 form the set {0, 3, —3, 6, —6, 9, —9, 12, —12 [illegible].}. The dots inside the brace indicate that there a [illegible] more members of the set besides those that are listed A characteristic of all integral multiples of 3 is that whe we divide them by 3, the remainder is 0.

The integers which are not multiples of 3 can be divide into two families. One family consists of those that are 1 more than an integral multiple of 3. For example, 7 = 6 + 1, and —5 = —6 + 1, so 7 and —5 both belong to th family. Members of this family have the characteristic th when we divide them by 3, the remainder is 1. The othe family consists of those integers that are 2 more than a integral multiple of 3. For example, 8 = 6 + 2, and —4 = —6 + 2, so 8 and —4 belong to this family. Members of thi family have the characteristic that when we divide them b 3, the remainder is 2. All integers therefore fall into one o three classes, depending on whether the remainder, when w divide by 3, is 0, 1, or 2. We call these classes *residue class* modulo 3. For convenience in talking about them, we sha designate each class by the remainder which is characteristi of it. Here, then, are the three classes:

$$\text{0 class: } \left\{\begin{matrix} 0, & 3, & 6, & 9, & 12, & \ldots\ldots \\ & -3, & -6, & -9, & -12, & \ldots\ldots \end{matrix}\right\}$$

$$\text{1 class: } \left\{\begin{matrix} 1, & 4, & 7, & 10, & 13, & \ldots \\ -2, & -5, & -8, & -11, & -14, & \ldots \end{matrix}\right\}$$

$$2 \text{ class: } \begin{Bmatrix} 2, & 5, & 8, & 11, & 14, \ldots \\ -1, & -4, & -7, & -10, & -13, \ldots \end{Bmatrix}$$

The 0 class has some interesting properties. Notice first that if we add any two members of the 0 class, the sum is itself a member of the 0 class. For example, $3 + (-6) = -3$. This is not true of the other two classes. For example, although 1 and 4 belong to the 1 class, their sum, $1 + 4 = 5$, which is not in the 1 class. Similarly, although 2 and 5 belong to the 2 class, their sum, $2 + 5 = 7$, which is not in the 2 class. Therefore, addition is a binary operation for the 0 class, but not for the 1 class or the 2 class. (Recall that a binary operation is defined for a set only by a mapping that associates with each pair of elements in the set a unique element *in the same set.*) Since addition is associative for the whole system of integers, it is surely associative when used with the members of the 0 class alone.

Notice, secondly, that the 0 class contains the identity element for addition, namely, 0. Moreover, for every element in the 0 class, its negative also belongs to the class. Therefore the 0 class satisfies all the requirements for being a group in its own right. In fact, since addition is commutative, it is even an abelian group. Because it is a group that is a subset of the larger group of integers, it is called a subgroup of the group of integers. The other two residue classes modulo 3 cannot qualify as subgroups with respect to addition, because, as we have seen, addition is not even a binary operation for them.

We can say even more about the 0 class. If we multiply any two members of the 0 class, the product is itself a member of the 0 class. So that multiplication is also a binary operation for this class. Moreover, it obeys the distributive law, so the 0 class satisfies all the requirements for being a ring in its own right. It is called a subring of the ring of integers.

We can say still more about the 0 class. If we multiply any member of the 0 class by *any integer*, whether it is in the class or not, the product is a member of the 0 class. For

example, $4 \cdot 3 = 12$, which is in the 0 class, although 4
not. A subring which has this property is called an *ide*
so the 0 class modulo 3 is an ideal in the ring of intege

The 0 class, considered as a ring, differs in one very i
portant way from the ring of integers. It does not conta
the number 1, which is the unity element for multiplic
tion. So we see it is possible for a ring to be with or wit
out a unity element.

The Arithmetic of Residue Classes

We are now going to construct a new number syst
hose elements are the residue classes modulo 3. To do
[illegible] ust define what we mean by addition and multiplic
f these classes. We define addition as follows: To a
two residue classes, pick one member from each class, a
add these members. The class to which the sum belongs w
be called the sum of those residue classes. For example, to a
the 1 class and the 2 class, we might add 1 (chosen from t
1 class) to 2 (chosen from the 2 class). The sum is 3, whi
belongs to the 0 class. Therefore the 1 class plus the
class equals the 0 class. Our definition allows us a free cho
in picking the numbers in each class that we will add. Ho
ever, it makes no difference which ones we choose, becau
all members of the same class have the same remaind
when divided by 3. Adding a representative from each
two classes is like adding their remainders, and the resu
is the same no matter which representative is chosen.

We can prove this fact for the illustration being giv
in this way: Any number in the 1 class is of the fo
$3a+1$. Any number in the 2 class is of the form $3b+$
When we add them, we get $3a+1+3b+2=3a+3b$
$3=3(a+b+1)$, that is a multiple of 3, which is in the
class. For the sake of brevity, let us now drop the word cl
from the name of each class, and simply refer to them
0, 1, and 2. Then we get the following addition table
residue classes modulo 3:

+	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

We define multiplication in the same way: To find the product of two residue classes, multiply any member of one class by any member of the other. The class to which the product belongs is the product of the classes. For example, to multiply the 2 class by the 2 class, multiply 2 (chosen from the 2 class) by 5 (chosen from the 2 class). The product is 10, which belongs to the 1 class. So the 2 class times the 2 class equals the 1 class. As in the case of addition, it makes no difference which member of each class is chosen for carrying out the operation. Here, too, we drop the word class from the name of each class, and record the products in the following multiplication table for residue classes modulo 3:

·	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

With these tables for addition and multiplication, the residue classes modulo 3 obey the five laws. Therefore they make up a number system consisting of only 3 elements. This number system is also both a group and a ring. It is a group with respect to addition, because 1) it contains a zero element, viz., the 0 class, and 2) for every element in the system, there is a negative in the system, too. The 0 class is a zero element, because, as the addition table sho that class added to any other class leaves it unchang $0+0=0$, $0+1=1$, and $0+2=2$. The negative of 0 0. The negative of 1 is 2, and vice versa, because $1+2$ $2+1=0$. The group is abelian, because the addition ope tion is commutative. The system is also a ring, because, b sides being an abelian group, it has a multiplication ope tion which is distributive with respect to addition. For e ple, $2(1+2)=2(0)=0$, and $2\cdot 1+2\cdot 2=2+1=$ Therefore $2(1+2)=2\cdot 1+2\cdot 2$.

Quotient Groups

The group that we have just observed contained as it nts a subgroup of the system of integers (the 0 class) rtain companion sets, the 1 class and the 2 clas The companion sets can be obtained from the subgrou by adding to each member of the subgroup some eleme that is not in it. For example, by adding 1 to each of th elements in the 0 class, we get all the elements of the class. In fact, if we add any single member of the 1 cla to each of the elements in the 0 class, we get all the me bers of the 1 class. Similarly, if we add any member of th 2 class to each of the elements of the 0 class, we get a the members of the 2 class. Companion classes that hav this relationship to a subgroup are called its *cosets*. Whe a subgroup and its cosets are themselves elements of group, the group they form is called a *quotient group*.

If we look back to the table for multiplication of the sym metries of a triangle (page 59), we shall find another exam ple of a subgroup, its cosets, and a quotient group th they form. Here, of course, the group operation will be * instead of +. Notice, first, that when I, P, and Q are mult plied among themselves, the product is always I, or P, or Q Therefore the operation * is a binary operation for the sub set $\{I, P, Q\}$. The identity element I is in this subset, an the inverse of every element in the subset is also in th subset. In fact, the inverse of I is I, the inverse of P is Q and the inverse of Q is P. Therefore the subset $\{I, P, Q$

forms a subgroup with the operation $*$. Its coset is the set $\{R, S, T\}$.

Now let us assign a short name to each of these sets, by writing $G = \{I, P, Q\}$, and $H = \{R, S, T\}$. We can define an operation $*$ on the system whose elements are G and H as follows. To multiply one set by another, pick an element from each set, and multiply them. For example, take P (from G) $*$ S (from H). The product is T, which is in H. The set to which the product belongs will be called the product of the sets. Therefore $G * H = H$. As in the case of residue classes, it turns out that the product comes out the same no matter which representative we choose from each class to carry out the operation. The products we get [illegible] be summarized in this multiplication table:

$*$	G	H
G	G	H
H	H	G

In this system of two elements, G is an identity element for the operation $*$, because $G * G = G$, and $G * H = H * G = H$. Also, there is an inverse for every element. In fact, each element is its own inverse, because $G * G = G$, and $H * H = G$, which is the identity element. Therefore the subgroup G and the coset H form a quotient group of the group of symmetries of the triangle. In this case, we can see why it is called a quotient group. The group of symmetries has 6 elements, the subgroup G has 3 elements, and the quotient group has $6 \div 3 = 2$ elements. Not every subgroup, together with its cosets, will form a quotient group. Galois was the first to identify the special kind of subgroups that have this property. They are known as normal subgroups. In an abelian group all subgroups are normal.

We saw that the residue classes modulo 3 form a ring as well as a group. It is an example of a *quotient ring*. Just as in a group, a special kind of subgroup with cosets forms a quotient group, in a ring, a special kind of ideal with i cosets form a quotient ring. In a ring in which the multipl cation is commutative, all ideals have this property.

Residue Classes Modulo 6

We formed the residue classes modulo 3 by classifyi the integers by the remainder you get when you divide by using for this purpose only positive remainders that are le than 3, viz., 0, 1 and 2. If we use any other integer as diviso we can divide the system of integers into residue class modulo that integer in the same way. For example, when a integer is divided by 6, the remainder may be 0, 1, 2, 3, [illegible] 5. This gives us six residue classes modulo 6. As we di [illegible] case of residue classes modulo 3, let us use the remainder associated with each class as the name of th class. If we define addition and multiplication of residu classes the same way we did before, we get these additio and multiplication tables:

+	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

·	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

With these tables, the residue classes modulo 6 form ring. This ring has a peculiar feature that it does not sha with the ring of integers. In the ring of integers, the onl time a product of two elements equals 0 is when one of th multipliers is itself 0. On the other hand, in the ring o residue classes modulo 6, $2 \cdot 3 = 0$, and neither 2 nor 3 i itself zero. Non-zero elements, like 2 and 3, which when multiplied give a zero product are called *zero divisors*. The ring of integers has no zero divisors, but the ring of residue classes modulo 6 does have zero divisors.

The absence of zero divisors in the ring of integers is the basis for one of the very important rules we all learned in elementary high school algebra, the cancellation law of multiplication. We learned that if $2 \cdot x = 2 \cdot 3$, we may cancel the 2 on both sides of the equation, and conclude that $x = 3$. The argument that proves this is correct proceeds as follows: Add the negative of $2 \cdot 3$ to both sides of the equation. This gives us $2 \cdot x - 2 \cdot 3 = 0$. By using the distributive law, we get $2 \cdot x - 2 \cdot 3 = 2\ (x - 3)$. Therefore $2 \cdot (x - 3) = 0$. This statement tells us that the product of two integers equals 0. This can happen only if one of the multipliers is 0. Since 2 is not 0, the other multiplier, $x - 3$, must be 0. Therefore x has to be 3. In the ring of residue classes modulo 6, this whole argument breaks down, because there *are* zero divisors in the system. That is, a product can be equal to 0, without either of the multipliers being equal to 0. Consequently, the cancellation law of multiplication is not obeyed in this system. In fact, in this system, if $2 \cdot x = 2$, we cannot conclude that $x = 1$. The multiplication table shows that while $x = 1$ is one possible solution to the equation, $x = 4$ is another solution, because $2 \cdot 4 = 2$.

Mapping Group into Group

We can set up a mapping of the group of integers into the group of residue classes modulo 3 by assigning to each integer in the system of integers the residue class that it belongs to, as shown in the table:

$$\begin{array}{c}
0 \rightarrow 0 \\
1 \rightarrow 1 \\
2 \rightarrow 2 \\
3 \rightarrow 0 \\
4 \rightarrow 1 \\
5 \rightarrow 2 \\
6 \rightarrow 0
\end{array}$$

By this mapping, 0, 3, 6, 9, and 12, for example, have the 0 class as their image; 1, 4, 7, 10, and 13 have the 1 class as their image; and 2, 5, 8, and 14 have the 2 class as their image. The mapping is clearly a many-to-one mapping. It has the interesting property of *preserving the operation* that is defined for the group. That is, the image of a sum is the sum of the images. For example, the image of 3 is 0, and the image of 4 is 1. The sum of 3 and 4 is 7, and its image is 1, which is the sum of the images 0 and 1. A mapping like this, of one group into another, which preserves the group operation is called a *group homomorphism*. Where it is a many-to-one mapping it is like collapsing or [illegible]scoping the group to make it fit into a smaller one. If we [illegible]r the system of integers and the system of residue classes modulo 3 as rings, the same mapping preserves not only the group operation of addition, but also the other ring operation, multiplication. For this reason, it is also an example of a *ring homomorphism* (one which preserves the operations in the ring).

When a group homomorphism of one group into another is a one-to-one correspondence, then it is a group *isomorphism*. Each element in one group is then paired off with one and only one element in the other. In that case, we say that the two groups are isomorphic to each other, or have the same structure. Two groups that are isomorphic to each other are really the same group structure dressed up in different clothes. We have an example of isomorphic groups in the group of residue classes modulo 3, and the subgroup $\{I, P, Q\}$ of the group of symmetries of the triangle. A mapping which matches them one-to-one, and preserves the group operation is shown below:

$$\begin{aligned} 0 &\longleftrightarrow I \\ 1 &\longleftrightarrow P \\ 2 &\longleftrightarrow Q \\ + &\longleftrightarrow * \end{aligned}$$

If we take any true statement in one system, such as $1+2=0$, and replace each symbol by its image under the mapping, we get a true statement in the other system. In the example shown, we get $P * Q = I$. In other words, the symbols used in the two systems are like two different languages that may be used for expressing the same ideas. The isomorphism printed above is the dictionary that allows us to translate from one language into the other.

DO IT YOURSELF

1. Using the definition of multiplication of integers given on page 48, prove that multiplication of integers is commutative by showing that

$$(a \sim b) \cdot (c \sim d) = (c \sim d) \cdot (a \sim b).$$

2. Using $(0 \sim a)$ and $(0 \sim b)$ to represent any two negative integers, prove that the sum of two negative integers is a negative integer.
3. Using $(0 \sim a)$ to represent any negative integer, and $(b \sim 0)$ to represent any positive integer, prove that the product of a negative integer and a positive integer is a negative integer.
4. A rearrangement of the numbers 1, 2, 3 is called a *permutation* of these numbers. Each rearrangement has the effect of replacing one number by another. For example, if the arrangement 123 is changed to 312, 1 is replaced by 3, 2 is replaced by 1, and 3 is replaced by 2. This permutation can be represented as a mapping:

$$\begin{gathered} 1 \rightarrow 3 \\ 2 \rightarrow 1 \\ 3 \rightarrow 2 \end{gathered}$$

There are six possible permutations of three numbers. Call them *I*, *A*, *B*, *C*, *D*, *E* as follows:

I	A	B
$1 \rightarrow 1$	$1 \rightarrow 2$	$1 \rightarrow 3$
$2 \rightarrow 2$	$2 \rightarrow 3$	$2 \rightarrow 1$
$3 \rightarrow 3$	$3 \rightarrow 1$	$3 \rightarrow 2$

C	D	E
$1 \to 1$	$1 \to 3$	$1 \to 2$
$2 \to 3$	$2 \to 2$	$2 \to 1$
$3 \to 2$	$3 \to 1$	$3 \to 3$

Define the product of two permutations $X * Y$ as the result of performing Y first and X afterwards on the result of Y. How a product is identified is shown in the following example: To find $A * B$:

B: A:		I:
$1 \to 3 \to 1$		$1 \to 1$
$2 \to 1 \to 2$	Result:	$2 \to 2$
$3 \to 2 \to 3$		$3 \to 3$

Therefore $A * B = I$.

a) Construct the multiplication table for the permutations I, A, B, C, D, E.

b) Prove that they form a group with the operation $*$.

c) Identify the subgroups of this group.

d) Show that this group of permutations is isomorphic to the group of symmetries of the triangle.

5. Use the associated remainders, 0, 1, 2, 3, 4 as the names of the residue classes modulo 5. Use the definitions of addition and multiplication given on pages 64-5 to construct addition and multiplication tables for these residue classes.

Use the tables to verify that $2 \cdot (1+2) = 2 \cdot 1 + 2 \cdot 2$ in this system. Does this system have zero divisors?

6. Show that the set of all even integers is a subgroup of the group of integers with respect to the operation $+$. Show that it is a subring of the ring of integers. Show that it is an ideal of the ring of integers. (See definitions, pages 63-4.)

CHAPTER IV

Numbers for Measuring

Another Defect to Overcome

THE natural number system has the defect that subtraction is not always possible in that system. To overcome this defect, we constructed the system of integers, an enlarged number system that includes the natural numbers, and in which subtraction *is* always possible. In this chapter we undertake another extension of the number system for a similar purpose. We find that the system of integers has the defect that *division* is not always possible within the system. To overcome this defect, we shall construct an enlarged number system that includes the integers, and in which division *is* (almost) always possible. The word "almost" has to be included in the statement of our goal, because, in the enlarged system, there will still be one number whose use as a divisor will be forbidden.

Just as subtraction of natural numbers was defined in terms of addition, division of integers may be defined in terms of multiplication. The symbol $\frac{-6}{2}$ really asks us the question, "What integer multiplied by 2 gives -6 as the product?" Since the answer to the question is -3, we say $\frac{-6}{2} = -3$. We call the symbol $\frac{-6}{2}$ the *quotient* of -6 and 2. We also refer to it as a *fraction*, and in this case it has meaning as another symbol for the integer -3. The question asked by the fraction $\frac{-6}{2}$ can also be written in the form of an equation, $2 \cdot x = -6$, and the answer to the question is the solution to the equation.

However, some fractions ask us a question that we cannot answer in the system of integers. For example, the fraction $\frac{2}{3}$ asks the question, "What number, when multiplied by 3, gives 2 as the product?" In the system of integers there isn't any such number. So, in the system of integers, the fraction $\frac{2}{3}$ has no meaning, and the equation $3 \cdot x = 2$ has no solution. This situation offers us a challenge similar to the one we faced in the last chapter. Can we build a number system in which division is always possible for any pair of numbers, so that $\frac{a}{b}$ always has a meaning, and $b \cdot x = a$ always has a solution?

Zero Is an Exception

We find that we can, provided that we agree not to use zero as a divisor. We can see the reason for this exclusion if we try to answer the question that is asked by a fraction that has 0 in the denominator. If the numerator *is not 0*, as in the fraction $\frac{2}{0}$, the fraction asks, "What number multiplied by 0 gives 2 as the product?" We hope to make the enlarged number system a ring. And in a ring, as we saw on page 61, zero times any other number gives a product equal to zero. Then the answer to the question will have to be "No number." On the other hand, if the numerator of the fraction *is 0*, the situation is even worse. Then the fraction $\frac{0}{0}$ asks, "What number multiplied by 0 will give 0 as a product?" The answer to this question would be, "any number." In fact, even in the system of integers, this would be the answer. We would like the fractions in our expanded number system not to overreach themselves. We want each fraction to stand for *one and only one number*. Since the fraction $\frac{2}{0}$ stands for no number, and the fraction $\frac{0}{0}$ stands for too many numbers, we exclude them as legitimate fractions. So, from now on, whenever we talk about a fraction, it will be understood that the denominator may not be 0.

Families of Fractions

To construct the new number system, we use the same device that was employed in the last chapter. To be sure that the question asked by the fraction $\frac{a}{b}$ (where b is not 0) always has an answer, we shall let the question be its own answer. We shall let each fraction represent a number in the new system. However, the numbers in this system will not be single fractions. Just as each number in the system of integers is a *family of differences* of natural numbers, each number in the system we are now constructing will be a *family of fractions or quotients* of integers. This is made necessary by the fact that, even among fractions that have a meaning in the system of integers, many fractions can represent the same number. For example, $\frac{3}{1}$, $\frac{6}{2}$, $\frac{9}{3}$, $\frac{12}{4}$, and many others represent the number 3, so we shall have to put them together into one family. In fact, we can obtain from this example the criterion we shall use for deciding when two fractions belong to the same family. Notice that $\frac{6}{2}$ and $\frac{9}{3}$, which belong to the same family, have the property that the numerator of each times the denominator of the other gives the same product, that is, $6 \cdot 3 = 9 \cdot 2$.

We are now ready to start building up the new number system. First, we take all possible ordered pairs of integers, such as 6 and 2, 5 and 7, -2 and 3, or -3 and -9, in which the second number in the pair is not zero. Then we write the "quotients" of the numbers in the pair, using the first number as numerator. So we now have a collection of fractions or quotients like $\frac{6}{2}$, $\frac{5}{7}$, $\frac{-2}{3}$, and $\frac{-3}{-9}$. Next, we associate with each fraction a family of fractions according to the following rule: The family of fractions belonging to $\frac{a}{b}$, where b is not zero, consists of all those fractions $\frac{u}{v}$ for which $a \cdot v = u \cdot b$. To designate the family, we shall write the fraction $\frac{a}{b}$ inside parentheses. Thus, $\left(\frac{a}{b}\right)$ means the family of fractions belonging to $\frac{a}{b}$. We call such a family of fractions a *rational* mber.

The rational numbers have some properties analogous to those of integers. First, a fraction belongs to its own family. Secondly, if one of two fractions belongs to the family of the other, then they have the same family. Because of these properties, each fraction belongs to one and only one rational number. A rational number can be represented by putting on display inside parentheses any one of the fractions that belongs to it. For example, since $\frac{9}{3}$ belongs to the rational number $\left(\frac{6}{2}\right)$, $\left(\frac{6}{2}\right) = \left(\frac{9}{3}\right)$. Finally, the criterion for membership in a rational number also serves as a test for equality of rational numbers. That is, the rational numbers $\left(\frac{a}{b}\right)$ and $\left(\frac{c}{d}\right)$ are equal if and only if $a \cdot d = c \cdot b$. For example, we know that $\left(\frac{6}{2}\right) = \left(\frac{9}{3}\right)$ because $6 \cdot 3 = 9 \cdot 2$.

Addition and Multiplication of Rational Numbers

Addition and multiplication of rational numbers are defined by the following equations:

$$\left(\frac{a}{b}\right) + \left(\frac{c}{d}\right) = \left(\frac{a \cdot d + b \cdot c}{b \cdot d}\right)$$

$$\left(\frac{a}{b}\right) \cdot \left(\frac{c}{d}\right) = \left(\frac{a \cdot c}{b \cdot d}\right)$$

For example, $\left(\frac{2}{3}\right) + \left(\frac{3}{5}\right) = \left(\frac{2 \cdot 5 + 3 \cdot 3}{3 \cdot 5}\right) = \left(\frac{19}{15}\right)$,

and $\left(\frac{2}{3}\right) \; \left(\frac{3}{5}\right) = \left(\frac{2 \cdot 3}{3 \cdot 5}\right) = \left(\frac{6}{15}\right)$.

To carry out the addition or multiplication of rational numbers, we apply these definitions to whatever fractions are on display inside the parentheses to represent the rational numbers. Since each rational number may be represented by any one of its members, the addition or multiplication may be carried out in many ways. The definitions make sense only if the result comes out the same no matter which member of a rational number is used to represent it. This turns out to be so. For example, we just obtained as the sum of $\left(\frac{2}{3}\right)$ and $\left(\frac{3}{5}\right)$ the rational number $\left(\frac{19}{15}\right)$. But $\left(\frac{2}{3}\right) = \left(\frac{4}{6}\right)$, since $2 \cdot 6 = 4 \cdot 3$; and $\left(\frac{3}{5}\right) = \left(\frac{12}{20}\right)$, since $3 \cdot 20 = 12 \cdot 5$. So we can add $\left(\frac{2}{3}\right)$ and $\left(\frac{3}{5}\right)$ by applying the definition to $\left(\frac{4}{6}\right)$ and $\left(\frac{12}{20}\right)$. We get the sum $\left(\frac{4 \cdot 20 + 6 \cdot 12}{6 \cdot 20}\right) = \left(\frac{152}{120}\right)$. But this is the same answer we got before, because $152 \cdot 15 = 19 \cdot 120$, showing that $\left(\frac{152}{120}\right) = \left(\frac{19}{15}\right)$.

With addition and multiplication defined in this way, the rational numbers form a number system, because they obey the five laws. We prove this fact here only for the commutative law of addition:

$$\left(\frac{a}{b}\right) + \left(\frac{c}{d}\right) = \left(\frac{a \cdot d + b \cdot c}{b \cdot d}\right)$$

$$\left(\frac{c}{d}\right) + \left(\frac{a}{b}\right) = \left(\frac{c \cdot b + d \cdot a}{d \cdot b}\right)$$

But the a, b, c, and d appearing in these symbols are integers, and integers obey the commutative laws for addition and multiplication. Therefore $b \cdot d = d \cdot b$, showing that the two results have the same denominator; and $a \cdot d + b \cdot c = c \cdot b + d \cdot a$, showing that the two results have the same numerator. Therefore the two sums are the same and $\left(\frac{a}{b}\right) + \left(\frac{c}{d}\right) = \left(\frac{c}{d}\right) + \left(\frac{a}{b}\right)$. A similar proof can be given for each of the other four laws.

Zero, Unity and Negatives

Like the natural number system and the system of in-ers, the rational number system has a zero element and a unity element. The rational number $\left(\frac{0}{1}\right)$ is the zero element, because $\left(\frac{0}{1}\right) + \left(\frac{x}{y}\right) = \left(\frac{0 \cdot y + 1 \cdot x}{1 \cdot y}\right) = \left(\frac{0 + x}{y}\right) = \left(\frac{x}{y}\right)$. The zero element can also be written in the form $\left(\frac{0}{b}\right)$, where b is any integer different from zero. This is proved by the test for equality of rational numbers, because $0 \cdot 1 = 0 \cdot b$. Where there is no danger of confusion with the zero of the system of integers, we use the symbol 0 for the zero element of the rational numbers, too.

The rational number $\left(\frac{1}{1}\right)$ is the unity element, because $\left(\frac{1}{1}\right) \cdot \left(\frac{x}{y}\right) = \left(\frac{1 \cdot x}{1 \cdot y}\right) = \left(\frac{x}{y}\right)$. The unity element can also be written in the form $\left(\frac{b}{b}\right)$, where b is any integer different from zero. This is proved by the fact that $1 \cdot b = b \cdot 1$. So $\left(\frac{2}{2}\right)$, $\left(\frac{3}{3}\right)$, $\left(\frac{4}{4}\right)$, and $\left(\frac{-5}{-5}\right)$ are all legitimate ways of representing the unity element. A rational number is the unity element if it is represented by a fraction whose numerator and denominator are equal. Where there is no danger of confusion with the unity element of the system of integers, we let the symbol 1 stand for the unity element of the rational numbers, too.

The system of integers has one important property that the natural number system does not have: every number in the system has a negative. The system of rational numbers has this property, too. In fact, if $\left(\frac{a}{b}\right)$ is any rational number, then $\left(\frac{-a}{b}\right)$ is its negative. To prove this fact, we must show that their sum is the zero element:

$\left(\frac{a}{b}\right) + \left(\frac{-a}{b}\right) = \left(\frac{a \cdot b + b \cdot (-a)}{b \cdot b}\right)$. The numerator of the fraction on display for the sum is $a \cdot b + b \cdot (-a)$. By the commutative law for multiplication of integers, $a \cdot b$ can be replaced by $b \cdot a$, so the numerator is equal to $b \cdot a + b \cdot (-a)$. By the distributive law for integers, this sum is equal to $b \cdot (a + (-a))$. But $a + (-a) = 0$, so we finally have that the numerator is equal to $b \cdot 0 = 0$. Therefore $\left(\frac{a}{b}\right) + \left(\frac{-a}{b}\right) = \left(\frac{0}{b \cdot b}\right)$ the zero element in the rational number system.

The Reciprocal of a Rational Number

The negative of a number was defined in terms of addition: one number is the negative of another if the sum of the two numbers is zero. The analogous concept in relation to multiplication is that of the *reciprocal*. One number is called the reciprocal of another if the product of the two numbers is the unity element. In the natural number system, 1 is the only number that has a reciprocal. In fact, it is its own reciprocal, since $1 \cdot 1 = 1$. In the system of integers, there are only two numbers that have reciprocals. They are 1 and −1. Each of them is its own reciprocal, because $1 \cdot 1 = 1$, and $(-1) \cdot (-1) = 1$. But in the rational number system, the existence of reciprocals becomes the rule rather than the exception. *Every rational number except the zero element in that system has a reciprocal.* In fact, if $\left(\frac{a}{b}\right)$

is any rational number that is not the zero element, then a is not 0. Consequently, $\left(\frac{b}{a}\right)$ is also a rational number, and it is the reciprocal of $\left(\frac{a}{b}\right)$. To prove that it is, we multiply them. The product $\left(\frac{a}{b}\right) \cdot \left(\frac{b}{a}\right) = \left(\frac{a \cdot b}{b \cdot a}\right)$ = the unity element, since the numerator and the denominator of the fraction on display are equal.

Since every rational number except the zero element has a reciprocal, it is convenient to introduce a special symbol to mean "the reciprocal of." If A stands for any rational number that is different from zero, we write $\frac{1}{A}$ for the reciprocal of A. Then, by the definition of reciprocal, we know that $A \cdot \frac{1}{A} = \frac{1}{A} \cdot A = 1$.

We undertook the construction of the system of rational numbers for the purpose of finding a number system in which the equation $B \cdot X = A$ always has a solution, as long as B is not zero. We can now show that we have achieved our purpose. If B and A are rational numbers, and B is not zero, then B has a reciprocal, $\frac{1}{B}$. If we multiply both sides of the equation by $\frac{1}{B}$ we get $\frac{1}{B} \cdot B \cdot X = \frac{1}{B} \cdot A$. But $\frac{1}{B} \cdot B = 1$, so we have $1 \cdot X = \frac{1}{B} \cdot A$. But $1 \cdot X = X$, because of the characteristic property of a unity element. Therefore we have found a solution to the equation. In fact, we know that X must be equal to $\frac{1}{B} \cdot A$. Since solving the equation $B \cdot X = A$ is another way of saying divide A by B, our result suggests a definition of division that is appropriate for rational numbers. To *divide* by a rational number means to multiply by its reciprocal. In the system of rational numbers, division is always possible as long as the divisor is different from 0.

We Still Have the Integers

By constructing the rational number system, we have gained something, in that every number in it except 0 has a reciprocal. At the same time, we have lost nothing, because we still have the original system of integers hidden within the rational number system in disguise. That is, there is a subset of the rational number system that is isomorphic to the system of integers, and therefore can take its place for all practical purposes. This subset consists of all rational numbers of the form $\left(\frac{a}{1}\right)$, where a is any integer, positive, negative, or 0. In fact, suppose we set up a mapping which matches each integer a with the rational number $\left(\frac{a}{1}\right)$. This mapping turns out to be an isomorphism, because, under this mapping, the image of a product is the product of the images, and the image of a sum is the sum of the images. The proof of this fact is seen in the following equations obtained by merely applying the definitions of addition and multiplication of rational numbers to the numbers $\left(\frac{a}{1}\right)$ and $\left(\frac{b}{1}\right)$:

$$\left(\frac{a}{1}\right) + \left(\frac{b}{1}\right) = \left(\frac{a \cdot 1 + 1 \cdot b}{1 \cdot 1}\right) = \left(\frac{a + b}{1}\right)$$

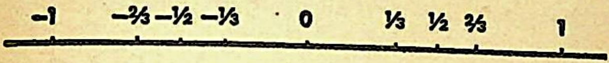
$$\left(\frac{a}{1}\right) \cdot \left(\frac{b}{1}\right) = \left(\frac{a \cdot b}{1 \cdot 1}\right) = \left(\frac{a \cdot b}{1}\right).$$

Because of this isomorphism, the rational number $\left(\frac{a}{1}\right)$ may be thought of as being the "same" as the integer a, and we use the symbol for the integer as an abbreviated notation for the corresponding rational number. So, from now on, we shall write 0 instead of $\left(\frac{0}{1}\right)$, and 1 instead of $\left(\frac{1}{1}\right)$, as already agreed. But we shall also write 2 instead of $\left(\frac{2}{1}\right)$, -2 instead of $\left(\frac{-2}{1}\right)$, and so on. We shall also drop the parentheses in writing a rational number, so that, when we write the fraction $\frac{1}{3}$, we shall mean the whole family of fractions of which $\frac{1}{3}$ is only a representative. With these conventions, we have the familiar notation for rational numbers that is used in everyday life. A rational number is called *positive* if it can be written as a fraction with positive numerator and denominator. It is called *negative*, if, when its denominator is positive, its numerator is negative.

The Rational Points on a Line

On page 29, we saw how we could represent the natural numbers as points on a line, spaced at equal intervals on one side of the point called 0. This procedure assigned numbers to only some of the points on the line. On page 54, we saw that we could represent the integers as points on a line, too. We equated the positive integers with the natural numbers already associated with points on one side of the 0. Then we placed the negative integers on the other side of the 0. In this way we assigned numbers to more of the points on the line. Now we can continue the process and assign numbers to many of the points that lie between those that represent the integers.

There is a point that divides the distance between 0 and 1 into two equal parts. We call that point $\frac{1}{2}$. There are two points that divide the distance between 0 and 1 into three equal parts. We call those points $\frac{1}{3}$ and $\frac{2}{3}$. By a similar procedure, we can put $-\frac{1}{2}$, $-\frac{1}{3}$, and $-\frac{2}{3}$ between 0 and −1.



This process can be extended so that we can find a point on the line for every rational number, positive or negative. The arrangement on the line makes it possible to talk about larger or smaller rational numbers in the same way that we could talk about larger or smaller integers. Of any two distinct rational numbers, the one that is further to the right is the larger one. Or, as in the case of integers, we can define the meaning of "larger" in this way: If a and b are rational numbers, a is larger than b if $a - b$ is positive.

The rational numbers are quite densely distributed over the whole line. Between any two points that represent rational numbers, there is at least another one that also represents a rational number. In fact, if a and b are two rational numbers, their average, $\frac{a+b}{2}$, is also a rational number, and lies between them. The fact that we can always find "in between" numbers makes the rational number system the appropriate one to use to represent measurements where subdivisions of the unit may be needed. By using as many "in between" numbers as we wish, we can refine a scale of measurement as much as we like, and make measurements as precise as the physical limitations of our equipment and our senses will allow us to.

Since the rational numbers are distributed densely over the line, we may guess that we now have a number assigned to every point on the line. But this guess turns out to be false, as we shall see. In fact, the lack of numbers for some points on the line is the next defect in the number system that we shall try to eliminate by expanding it once more.

The Rational Numbers Form a Field

The rational number system, like the system of integers, is an example of a group structure and a ring structure.

Addition of rational numbers is associative. There is an identity element for addition (the zero), and each element has an inverse with respect to addition (the negative). So the rational number system meets all the requirements for being a group with respect to addition. In fact, since addition of rational numbers is commutative, it is an abelian group, and we are conforming to custom by using a plus sign for the group operation. The operation of multiplication in the rational number system is associative, and it is also distributive with respect to addition. With these further properties, the rational number system meets the requirements for being a ring. In fact, since multiplication is commutative, it is a commutative ring. Moreover, it contains a unity element.

In the transition from integers to rational numbers, something new has been added. Every rational number except 0 has a reciprocal. But a reciprocal is simply an inverse with respect to the operation of multiplication. So the rational number system with zero omitted meets all the requirements for being a group with respect to multiplication. It therefore has a double group structure, one for addition, and one for multiplication. A system of this kind, that has a double group structure, is called a *field*. A field is defined as a ring in which a unity element exists, and which has a reciprocal for every element except zero. The presence of these reciprocals makes it possible to carry out division by any element except zero.

The way in which groups, rings, and fields differ from each other may be expressed briefly, though crudely, as follows: A group is a system in which we can perform addition and subtraction. A ring is a system in which we can perform addition, subtraction, and multiplication. A field is a system in which we can perform addition, subtraction, multiplication, and division, except that division by 0 is excluded.

Finite Fields

The rational number system is a field that contains an infinite number of elements. There are also fields that have only a finite number of elements. In fact, we have already encountered some in earlier sections of this book. In the last chapter, we found that the system of residue classes of integers modulo 3 has a ring structure. The elements in the ring are classes called 0, 1, and 2. The multiplication table in this ring is as follows:

·	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

We see from this table that every element in the ring except zero has a reciprocal. In fact, the reciprocal of 1 is 1, because $1 \cdot 1 = 1$; and the reciprocal of 2 is 2, because $2 \cdot 2 = 1$. So the system of residue classes of integers modulo 3 is a field.

If you did exercise 5 in the "Do It Yourself" section at the end of Chapter III, you have met another finite field. The residue classes of integers modulo 5 constitute a ring with this multiplication table:

·	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

This table shows that every element except zero in this system has a reciprocal. In fact, the reciprocal of 1 is 1; the reciprocal of 2 is 3; the reciprocal of 3 is 2; and the reciprocal of 4 is 4.

On the other hand, the residue classes of integers modulo

6 constitute a ring that is *not* a field. The multiplication table for this system, printed on page 68, shows that some of its elements which are not zero do not have a reciprocal. In fact, there is no element whose product with 2, 3, or 4 is equal to 1. So the elements 2, 3, and 4 in this system do not have reciprocals. The reason for this failure is that the number 6 has positive integer divisors besides itself and the number 1. These divisors are 2 and 3. Any integer which is also divisible by 2 or 3 belongs to a residue class that cannot have a reciprocal. Let us prove this fact for an integer divisible by 2.

Suppose that such an integer belongs to residue class a, and we multiply by any other residue class x. We shall show that the product $a \cdot x$ cannot be equal to 1, and so a cannot have a reciprocal. To identify the product of a and x, we follow the directions given on page 65. We pick any member of a and any member of x, multiply them, and then identify the residue class of the product. Identifying the residue class of the product means finding its remainder when it is divided by 6. As the representative of a let us use the member that we know is in it that is divisible by 2. Since it is divisible by 2 we may represent it as $2 \cdot m$, where m is some other integer. Let us designate by k the member we select from class x. Then the product of the two representatives of their classes is $2 \cdot m \cdot k$. Now we divide this product by 6, and obtain a quotient and a remainder, both of which are integers. Let us call the quotient q, and the remainder r. Then, from the fact that a dividend is equal to the divisor times the quotient plus the remainder, we can say that $2 \cdot m \cdot k = 6 \cdot q + r$. Therefore $r = 2 \cdot m \cdot k - 6 \cdot q$. By the distributive law, $r = 2 \cdot (m \cdot k - 3 \cdot q)$. That is, the remainder is divisible by 2. Therefore it cannot be 1. Therefore the residue class to which $2 \cdot m \cdot k$ belongs is not class 1. Therefore the product of class a and class x cannot be class 1, no matter what class x may be.

By means of a similar proof, it can be shown that, in general, the ring of residue classes modulo n is not a field if n has positive integer divisors besides itself and the number

1. An integer that has positive integer divisors other than itself and 1 is called *factorable*. An integer that is not factorable is called prime. It can also be shown that if n *is* a prime integer, the ring of residue classes modulo n *is* a field. The integers 3 and 5 are both prime. That is why the ring of residue classes modulo 3 and the ring of residue classes modulo 5 both turned out to be fields.

No Zero Divisors

In the system of integers, we found that the cancellation law for multiplication was a consequence of the fact that the ring of integers has no zero divisors. Since this law is a great convenience in solving equations, it would be useful if it turned out to be true in the rational number system, too. Fortunately, it does, because the rational number system is a field, and *a field cannot have zero divisors*. The proof of this fact flows directly from the definition of zero divisors and the definition of a field. If a field did have zero divisors, there would be two elements in the field, say a and b, both not zero, but whose product is zero. So we could write $a \cdot b = 0$. Since a is not zero, and the system is a field, it has a reciprocal, $\frac{1}{a}$. Multiplying both sides by $\frac{1}{a}$, we get $\frac{1}{a} \cdot a \cdot b = \frac{1}{a} \cdot 0$. But $\frac{1}{a} \cdot a = 1$, and $\frac{1}{a} \cdot 0 = 0$, so we have $1 \cdot b = 0$, or $b = 0$, contradicting the assumption that a and b are both not zero. Therefore it is impossible for a field to have zero divisors.

Ideals in a Field

The special properties of a field also impose some restrictions on the ideals that it may contain. To get acquainted with these restrictions, let us first recall the definition of an ideal, and note some facts about the ideals of rings in general. We defined an ideal as a subring of a ring that has the property that if we multiply any member of the subring by any member of the ring, whether it is in the subring or not, the product turns out to be in the subring. For example, the set of all *even* integers is an ideal in the ring of integers because, first, it is a subring, and secondly, the product of an even integer and *any* integer is an even integer.

Every ring contains at least two ideals. One of them is the subset consisting of only one element, the 0 of the ring. The other is the entire ring itself.

To show that the subset that contains only the 0 element is an ideal, we check to see if it fits the definition of an ideal. Notice first that addition within this subset is associative and commutative because it is in the ring as a whole. Multiplication within the subset is associative, and is distributive with respect to addition because it is in the ring as a whole. We observe next that the subset $\{0\}$ contains the zero element, and also the negative of each of its elements, since 0 is its own negative. Moreover, $0 + 0 = 0$, so the sum of elements in the subset is in the subset. Therefore the subset meets all the requirements for being an abelian group. The product $0 \cdot 0 = 0$, and so is a member of the subset. And, since multiplication in it is associative and distributive with respect to addition, the subset meets all the requirements for being a subring. Now we observe that any element in the ring times 0 gives a product equal to 0, so that this product is also in the subring. This last property makes the subring an ideal. Similarly, checking the requirements one by one, we see that the original ring is one of its ideals.

The ideal that consists of the zero element alone is called the *zero ideal.* We shall also assign a special name to the original ring when viewed as one of its own ideals. We call it the *unit ideal.* The reason for this name is seen in the following considerations. Suppose the ring we are talking about has a unity element. (This is not true of all rings, as we saw on page 64, but it is true of all fields.) Let us take a close look at any ideal that contains the unity element 1. The characteristic property of an ideal is that when we multiply a member of the ideal by any member of the ring, the product is a member of the ideal. Let us, then, multiply 1 by any element x in the ring. The product, $x \cdot 1 = x$ is therefore in the ideal. In other words, every element in the ring is in the ideal. So, if an ideal contains the unity element, it contains all the elements of the ring, and therefore must be the whole ring. That is why the ideal that consists of the whole ring is called the unit ideal. We have in this fact and in this name a test for recognizing when an ideal is the unit ideal. To show that an ideal is the unit ideal of a ring, it suffices to show that the ideal contains the element 1.

Since a field is a ring, it has these two special ideals, the zero ideal and the unit ideal. Now we show that a field has no other ideals besides these two. Suppose we examine any ideal of a field. Since the ideal is a subring, it must contain the 0 element. If it contains no other elements, then it is the zero ideal. If it contains some other element, say b, then b is different from zero, and therefore has a reciprocal $\frac{1}{b}$ in the field. If we multiply the element b of the ideal by $\frac{1}{b}$, the product is a member of the ideal. But this product is 1. Therefore the ideal contains 1, and must be the unit ideal. This concludes the proof.

DO IT YOURSELF

1. Separate all integers into residue classes modulo 7, by putting into one class all the integers that have the same remainder when you divide by 7. By using the remainder as the name for the class associated with it, we get seven classes called 0, 1, 2, 3, 4, 5, and 6. Construct the multiplication table for these residue classes (see page 65). Verify from the table that each element in this system except 0 has a reciprocal. What are the reciprocals of 2, 3, 4, 5, and 6 in this system?
2. Construct the multiplication table for residue classes modulo 12. Which elements are zero divisors in this system?
3. Prove that multiplication of rational numbers is commutative.
4. Prove that addition of rational numbers is associative.

CHAPTER V

Filling Out the Line

More Questions to Be Answered

SO FAR we have expanded the number system twice, from natural numbers to integers, and then from integers to rational numbers. In each case, the purpose of the expansion was to have a number system in which a certain type of equation would always have a solution. The first type of equation we tried to solve was one that included only a single step of addition. This was the equation of the form $b + x = a$. The second type of equation we tried to solve included only a single step of multiplication. This was the equation of the form $b \cdot x = a$. It was natural to consider these equations first, because addition and multiplication are the operations that are built into the structure of a number system.

It is just as natural now to go beyond these simplest equations, and consider others that may include *both* operations, or may use an operation *more than once*. For example, we might examine equations like $2 \cdot x + 5 = 11$, or $3 \cdot (x + 2) = 7$ which involve both addition and multiplication. We could also consider an equation like $3 \cdot x \cdot x \cdot x - 5 \cdot x \cdot x = 2 \cdot x + 9$ where the operation of multiplication is repeated several times. Such equations, in which only addition and multiplication may be employed, are known as *algebraic equations*.

For a systematic survey of these equations, we first rewrite them in standard form. Wherever the unknown "x" is multiplied by itself several times, we use the abbreviated power notation, writing x^2 for $x \cdot x$, x^3 for $x \cdot x \cdot x$, and so on. We eliminate parentheses by using the distributive law.

For example, $3 \cdot (x + 2)$ can be replaced by $3 \cdot x + 6$. We reduce one side of the equation to 0 by adding negatives where necessary. For example, in the equation $x^2 - 3x = x + 6$, if we add $-x - 6$ to both sides of the equation, we get $x^2 - 3x - x - 6 = 0$. Finally, we combine like terms, and arrange all terms in descending powers of x, getting $x^2 - 4x - 6 = 0$.

Here are some typical equations, written in standard form, using numbers that belong to the rational number system:

$$\frac{2}{3} \cdot x - \frac{5}{2} = 0.$$

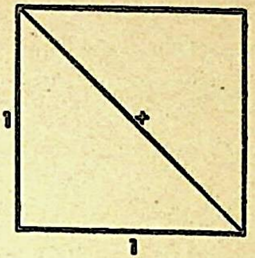
$$\frac{3}{7} \cdot x^2 + \frac{5}{3} \cdot x - \frac{1}{2} = 0$$

$$x^3 - 2 \cdot x^2 + \frac{1}{3} \cdot x - 5 = 0$$

The highest power of x that appears in an algebraic equation in standard form is called the *degree* of the equation. *In the rational number system, an equation of the first degree always has a solution.* The typical first degree equation has the form $a \cdot x + b = 0$, where a and b are rational numbers, and a is different from 0. To solve it, we take advantage of the fact that every rational number has a negative, and every rational number that is different from 0 has a reciprocal. First we add to both sides of the equation the number $-b$, which is the negative of b. This gives us $a \cdot x = -b$. Then we multiply both sides of the equation by $\frac{1}{a}$, which is the reciprocal of a, and we find that $x = \frac{1}{a} \cdot (-b)$. This result tells us that if there is a solution to the equation, it must be equal to $\frac{1}{a}(-b)$. We verify that it is indeed a solution by substituting into the equation. For example, to solve the equation $\frac{2}{3} \cdot x - \frac{5}{2} = 0$, first add $\frac{5}{2}$ to both sides of the equation. Then multiply both sides by $\frac{3}{2}$. The result asserts that if there is a solution, it must be equal to $\frac{3}{2} \cdot \frac{5}{2}$, or $\frac{15}{4}$. If we substitute $\frac{15}{4}$ for x in the equation, it then says $\frac{2}{3} \cdot \frac{15}{4} - \frac{5}{2} = 0$. Calculation shows that this statement is true, so that $\frac{15}{4}$ really is a solution to the equation.

However, we have less luck when we try to solve equations of the second degree. In the rational number system, we can solve some of them, but we cannot solve others. For example, we can solve without any difficulty the equation $x^2 - 1 = 0$. First add 1 to both sides and we get $x^2 = 1$. In this form, the equation asks, "What number, when multiplied by itself, gives 1 as the product?" The number 1 is obviously an answer to this question, because $1 \cdot 1 = 1$. In fact, the number -1 is also a good answer to the question, because $(-1) \cdot (-1) = 1$. So we have found two solutions to the equation. The equation $x^2 - 2 = 0$ looks as though it ought to be just as easy to solve, but it isn't. If we add 2 to both sides, we get $x^2 = 2$. In this form, the equation asks, "What number, when multiplied by itself, gives 2 as the product?" We may be tempted to say that the answer is $\sqrt{2}$, or the square root of 2. But this is really not an answer to the question. It is only a restatement of the question. When we write the symbol $\sqrt{2}$, it signifies, "that number (if it exists) which when multiplied by itself gives 2 as the product." But the question still remains, "Does it exist?" We shall soon see that it does not exist in the rational number system. We shall prove that there is no rational number whose square is equal to 2.

Before giving the proof, let us approach the question in another way. We find that the question we are trying to answer is one that comes up in a simple problem in geometry. Suppose we construct a square whose sides are one unit long, and then draw the diagonal of the square. The diagonal has a definite length. To calculate it, we use the Pythagorean



theorem that states that the square of the hypotenuse of a right triangle is equal to the sum of the squares of the legs. In this case the theorem leads to the equation $x^2 = 1^2 + 1^2$, or $x^2 = 2$, which is precisely the equation that we are trying to solve. So, when we prove that there is no rational number whose square is equal to 2, we shall be proving at the same time that no rational number can represent the length of the diagonal of a square whose side has length 1.

The proof makes use of some simple facts about rational numbers and integers that we shall take note of first. 1) Every rational number can be represented by a fraction that is "reduced to lowest terms." 2) If a fraction is reduced to lowest terms, then its numerator and denominator cannot both be even numbers. For, if they were, it would mean that the fraction can be reduced further by dividing numerator and denominator by 2. For example, $\frac{6}{8}$ is not in lowest terms. It can be reduced to $\frac{3}{4}$. 3) The square of an even integer is an even integer, and the square of an odd integer is an odd integer. For example, $6 \cdot 6 = 36$, which is an even integer; $7 \cdot 7 = 49$, which is an odd integer. From this fact, it follows that if the square of an integer is even, then the integer itself is even.

We prove that there is no rational number whose square is equal to 2 by showing that the assumption that there is one leads to a contradiction. Suppose there is a rational number whose square is equal to 2. Then it can be represented by a fraction that is reduced to lowest terms. Let $\frac{a}{b}$ represent this fraction, where a and b are both integers. Since the fraction is in lowest terms, *a and b are not both even integers.* Since the square of this fraction is supposed to be equal to 2, we can write $\frac{a^2}{b^2} = 2$. Multiplying both sides by b^2, we find that $a^2 = 2 \cdot b^2$. This equation tells us that a^2 is double the integer b^2, so a^2 is an even integer. But if a^2 is even, then a must be even, that is, it is double some other integer. If we call that other integer k, then $a = 2 \cdot k$. In that case, $a^2 = (2 \cdot k)(2 \cdot k) = 4 \cdot k^2$. If we substitute this expression for a^2 in the equation $a^2 = 2 \cdot b^2$, we get $4 \cdot k^2 = 2 \cdot b^2$. Dividing both sides by 2, we see that $2 \cdot k^2 = b^2$. In other words, b^2 is double the integer k^2, or b^2 is an even integer. But if b^2 is even, *then b must be even.* We began by observing that a and b are not both even, and end up by concluding that they are both even. We were led to this contradiction by the assumption that there is a rational number whose square is equal to 2. Therefore we are compelled to reject the assumption.

The proof given above is a very old one. It was first worked out about 2500 years ago by the Greek mathematician Pythagoras. The philosophers of Greece were so pleased to discover that there were lengths that could not be represented by rational numbers, that they celebrated the discovery, we are told, by sacrificing one hundred oxen to the gods.

In Chapter IV we represented the rational numbers as points on a line. In this representation, wherever a positive number is attached to a point, the number represents the distance of that point from 0, if we use the distance from 0 to 1 as the unit of length. Suppose, now, we measure out to the right of 0 a length equal to the length of the diagonal of a square whose side has length 1 (see diagram on page 102). In this way we locate a definite point whose distance from 0 is equal to the length of the diagonal. But we have just proved that there is no rational number that can represent this length. So we have found a point on the line that has no rational number attached to it. This discovery answers the question that we raised on page 83. The rational number system supplies numbers for some, but not all, of the points on the line. Even though the rational numbers are spread out densely over the line, there are gaps that they leave unfilled. If we want to have a number to represent every point on the line, we have to expand the number system again. We proceed to do so now, in order to fill the gaps.

Decimal Fractions

There are many ways of approaching the problem of filling the gaps between the rational numbers. We shall use here an approach that grows naturally out of the custom of writing numbers as "decimals." Decimals are more correctly described as decimal fractions. The decimal .2 is an abbreviated way of writing the fraction $\frac{2}{10}$. The decimal .23 is an abbreviated way of writing the fraction $\frac{23}{100}$. In each case, the denominator is understood to be a power of 10, and we identify the power by counting the number of digits to the right of the decimal point. Thus, since .235 has three digits after the decimal point, we know the denominator is 10^3 or 1000, and the decimal .235 represents the fraction $\frac{235}{1000}$. So we see that each decimal represents a fraction, and therefore is simply another way of writing some of the numbers in the rational number system. This observation immediately suggests the question, "Can every rational number be written as a decimal?" We are led to a rather interesting problem when we try to answer this question.

In elementary school we learned that we can convert a fraction into a decimal by using the process of long division. For example, to get the decimal equivalent of $\frac{1}{4}$, we divide 4 into 1, using the following form:

$$\begin{array}{r}
.25 \\
4\,\overline{)\,1.00} \\
\underline{8} \\
20 \\
\underline{20}
\end{array}$$

The long division terminates after two steps because in the last subtraction the remainder is zero. The conclusion is that $\frac{1}{4} = .25$. If we try the same procedure to get a decimal equivalent for the fraction $\frac{1}{3}$, we run into trouble. We arrange the work in the same way, as follows:

$$\begin{array}{r}
.3333 \\
3\,\overline{)\,1.0000} \\
\underline{9} \\
10 \\
\underline{9} \\
10 \\
\underline{9} \\
10 \\
\underline{9} \\
1
\end{array}$$

But, no matter how many steps we carry out, the division never comes to an end. After each subtraction there is a remainder of 1. So the fraction $\frac{1}{3}$ cannot be represented as a decimal with a finite number of digits.

The persistent reappearance of the remainder 1 tempts us to keep dividing. If we do so, we get a decimal that never ends. If we want to represent the fraction $\frac{1}{3}$ by a decimal at all, it will have to be an *infinite decimal,* that is, one that has an infinite number of digits after the decimal point. So, to decide whether every rational number can be represented as a decimal we have to investigate what meaning, if any, an infinite decimal can have.

To interpret the meaning of a finite decimal, we identify a numerator from the digits that we see, and a denominator by using the appropriate power of ten, indicated by the number of digits that appear after the decimal point. Then we put the two together in the form of a fraction. This method breaks down with an infinite decimal, so we have to try another approach. What we do is think of the infinite decimal as a sequence of finite decimals, made progressively longer by appending another digit at each step. In this view, the infinite decimal .333333 . . . . represents the infinite sequence of finite decimals .3, .33, .333, .3333, .33333, ... To see how this sequence is related to the fraction $\frac{1}{3}$, let us compare each of the numbers in the sequence with the fraction. Suppose, for example, we subtract .3 from $\frac{1}{3}$. Writing the decimal as a common fraction, we say $\frac{1}{3} - \frac{3}{10} = \frac{10}{30} - \frac{9}{30} = \frac{1}{30}$. This difference is rather small, and in many practical problems is small enough to be considered negligible. So we may use .3 as an approximation of the value of $\frac{1}{3}$. If we subtract .33 from $\frac{1}{3}$, we find the difference to be $\frac{1}{300}$. This difference is smaller than $\frac{1}{30}$, so .33 is a better approximation to the value of $\frac{1}{3}$ than .3 is. If we try each of the decimals in the sequence in turn, we get better and better approximations to the number $\frac{1}{3}$. The longer the decimal is, the better the approximation becomes, because the difference from $\frac{1}{3}$ gets smaller and smaller. To approximate means literally to come close. By taking more and more digits in the decimal we get a number that is closer and closer to $\frac{1}{3}$. In fact, we can get as close to $\frac{1}{3}$ as we please by simply taking a decimal that is long enough. We describe the situation by saying that the sequence of decimals approaches the number $\frac{1}{3}$ as a limit.

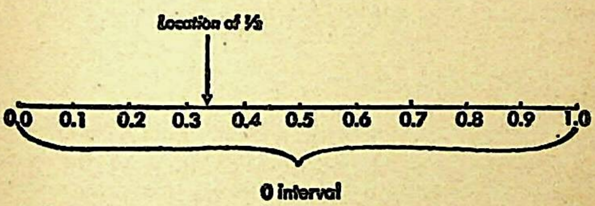
We now have the means of explaining how an infinite decimal may represent a number. An infinite decimal represents a number if the sequence of finite decimals obtained by taking more and more of its digits approaches that number as a limit. With this definition of the meaning of an infinite decimal, it may make as much sense as a finite decimal, and we can now answer the question, "Can every rational number be represented by a decimal?" The answer is, "Yes, by a finite or infinite decimal."

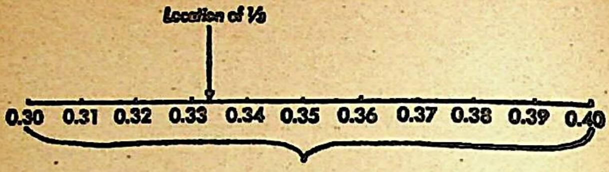
A Nest of Intervals

The meaning of an infinite decimal can also be expressed pictorially in terms of our representation of the rational numbers as points on a line. It can be interpreted as a description of the position of a number, or directions that may be followed in order to find it.

Suppose, for example, we want to describe where the number $\frac{1}{3}$ is located. Since it is a rational number, we can say first that it is somewhere on the line on which the rational numbers are represented. Then we try to specify where it is on the line. The integers on the line divide the line into intervals of unit length. We can assign a name to each interval by using the number that is attached to the end that is nearest to 0. We call the interval between 0 and 1 the "0 interval"; the interval between 1 and 2 the "1 interval," and so on. On the negative side of the line, we call the interval between 0 and −1 the "−0 interval," the interval between −1 and −2, the "−1 interval." Notice that 0 and −0 refer to different intervals.

Now we can be more specific about where the number $\frac{1}{3}$ is by saying that it is in the 0 interval. So we write down 0 as the first part of a chain of directions. The next step in the description is to say where it is in the 0 interval. For this purpose, we divide the 0 interval into ten equal parts, each with length .1. We label these intervals in order, starting with the one nearest to the 0, as 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9. Now we learn from long division that $\frac{1}{3}$ is more than 0.3, but less than 0.4, so it lies in the interval whose name is 0.3. The name of this interval incorporates both parts of the description we have given so far. It tells us that the number $\frac{1}{3}$ is in the 0 interval, and that within that interval it is in the sub-interval that has 0.3 as its left end. Now we narrow down its location once more. We divide the sub-interval into ten equal parts, each of width .01, and label them, in order, 0.30, 0.31, 0.32, 0.33, 0.34, 0.35, 0.36, 0.37, 0.38, and 0.39. Again we learn from long division that $\frac{1}{3}$ is more than 0.33, but less than 0.34. So it lies in the interval whose name is 0.33. The name of this interval is a three part description. It says that the number $\frac{1}{3}$ is in the 0 interval, and inside this interval it is in the 0.3 interval, and inside the latter interval it is in the 0.33 interval. Now we subdivide this interval, and continue the process indefinitely.





0.3 interval enlarged

Each time we pick out the sub-interval that contains the point. According to this scheme, the infinite decimal 0.33333 . . . . represents an infinite sequence of intervals, with the following characteristics: each successive interval lies inside the interval that precedes it in the sequence, and as we move along the sequence, the width of the intervals shrinks toward 0. We call such a sequence of intervals a *nest of intervals.*

In this instance, we chose each interval in the nest as an interval that contains the number $\frac{1}{3}$. So we are sure that the number $\frac{1}{3}$ lies inside every one of the infinite number of intervals in the nest. We can also be sure that no other number can be in there with it, because two different numbers cannot lie in the same nest. This is so, because two different numbers are separated by a distance that is greater than 0. Since the inner intervals in a nest shrink in width toward a width of 0, ultimately the innermost intervals are too narrow to span the distance between two separate points, and cannot enclose them both. For this reason, the nest of intervals represented by the infinite decimal 0.333 . . . : describes the number $\frac{1}{3}$ and no other number. *A nest of intervals represents a number if that number is attached to the only point that lies inside all the intervals of the nest.*

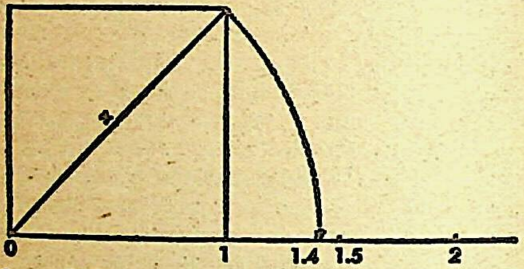
A Nest for Every Rational Number

By following the same procedure with any rational number, we can find a nest of intervals that contains the number, and the nest of intervals can be represented by an infinite decimal. In this way we can get an infinite decimal for every rational number. This is true even for numbers that are represented by finite decimals. In fact, in these cases, the number can be represented by two infinite decimals. Suppose, for example, we look for a nest of intervals and the associated infinite decimal to represent the number 2.3. First we notice that the number lies in the 2 interval which extends from 2 to 3. Now, when we subdivide this interval into ten equal parts, one of the points of division is actually 2.3. As a point of division, it belongs to *two* intervals. It is the right end of the interval whose name is 2.2. It is also the left end of the interval whose name is 2.3. So we may choose either one as the interval to which it belongs. After we have made that choice, we shall not be free to choose again. If we think of the number as belonging to 2.2, then forever after, as we subdivide the intervals, the number will always lie in the last subdivision. The infinite decimal in this case comes out 2.2999999 . . . . with an endless series of nines. If we think of the number as belonging to 2.3, then forever after, as we subdivide the intervals, the number will always lie in the first subdivision. Then the infinite decimal comes out 2.3000000 . . . . with an endless series of zeros. We get two names like this for every number that can be represented as a finite decimal, because a finite decimal turns up, sooner or later, as an endpoint of an interval in the nest. The two different ways of representing the number signify that such a number can be approached from two directions, from the left or from the right.

A Nest That Has No Number

We have answered in the affirmative the question, "Can every rational number be represented by a decimal?" We have found that it can be represented by an infinite decimal, and in some cases by two infinite decimals. In each case the infinite decimal represents a nest of intervals, and the rational number is attached to the only point that is inside the nest.

Now let us reverse the question and ask, "Does every infinite decimal represent a rational number?" Or, to ask the question in another form, "Does every nest of intervals defined by an infinite decimal have a rational number attached to the point inside the nest?" We can see immediately that the answer to this question is, "No." We found before that there are some points on the line that have no rational number attached. The point whose distance from 0 is equal to the length of the diagonal of a unit square is such a point. Its location is pointed out by the arrow in the diagram. However, we can follow the same procedure with this point as we did with the point that represents the fraction $\frac{1}{3}$. We can locate it precisely by means of a nest of intervals, and the associated infinite decimal. We observe first from the diagram that the point lies between 1 and 2 so the decimal begins with the number 1. Then we subdivide the interval between 1 and 2, and observe that the point lies between 1.4 and 1.5. In fact, we can verify this fact by arithmetic, by observing that $1.4 \times 1.4 = 1.96$, a number that is less than 2, while $1.5 \times 1.5 = 2.25$, a number that is more than 2. So, 1.4 is too small to be the length of the diagonal of the unit square, and 1.5 is too large. By continuing the process of subdivision, we get a nest of intervals and an infinite decimal that represents it. There is one and only one point inside this nest, the point whose distance from 0 is the length of the diagonal of the unit square. But there is no rational number attached to this point. So the infinite decimal that belongs to this nest does not represent a rational number.



The Real Number System

We can think of an infinite decimal as a question. It asks, "If you make the nest of intervals that my digits describe, what number will you find inside the nest?" In the rational number system this question does not always have an answer. So, once again, we ask, "Is there a larger number system in which such a question does always have an answer?" To construct such a number system, we use the same device we have already used twice before. We make the question its own answer. We build a new number system in which each element *is* an infinite decimal. We call this system the system of *real* numbers. In general, each separate infinite decimal is a separate number. But, we know from our experience with the rational number 2.3 that we shall have to provide for some exceptions. We must specify that whenever a decimal ends with an infinite chain of nines, as 2.99999 . . . does, the decimal obtained by replacing the chain of nines by a chain of zeros and adding 1 to the last digit before the chain will represent the same number. Thus 2.300000 . . . . and 2.299999 . . . . will represent the same number. Under this agreement, every real number is either a single infinite decimal, or a pair of infinite decimals either one of which may be used to represent the pair.

When we represented the rational numbers as points on a line, we found that there were gaps on the line, where no numbers were attached. Later we found that every point on the line can be located by an infinite decimal, and now we are converting every infinite decimal into a number. So, at one stroke, we are filling all the gaps on the line. The real number system gives us a number for every point on the line.

So far we have merely defined a collection of elements, each of which is an infinite decimal or a pair of decimals. To convert this collection into a number system we have to give it the structure that every number system must have. We have to define operations of addition and multiplication on these elements, and we must show that the operations obey the five laws. To define the operations, we lean on the smaller number system we already have as a crutch. Just as we used the natural numbers when we built the structure of the system of integers, and we used the integers when we built the structure of the system of rational numbers, we now use the rational numbers when we build the structure of the system of real numbers. We define addition and multiplication of real numbers as follows:

To add two infinite decimals, break up each decimal into a sequence of finite decimals, by using in succession, more and more of the digits in the decimal. These finite decimals represent rational numbers, and may be added by the rules for rational numbers. So, first add the first numbers in each sequence. The sum is a first approximation to your answer. Then add the second numbers in each sequence. This gives you a better approximation to your answer. As you proceed with the sequence of additions, you get a longer and longer finite decimal. The digits at the beginning of the decimal may fluctuate at first, but then settle down so that ultimately a fixed digit is defined for each decimal place. In this way a specific infinite decimal is obtained as the sum. To multiply infinite decimals, multiply the successive pairs of finite decimals in the same way.

For example, to add 0.2222222 . . . . and 0.8888888 . . . ., we proceed by these steps:

$$\begin{array}{r} 0.2 \\ +0.8 \\ \hline 1.0 \end{array} \quad \begin{array}{r} 0.22 \\ +0.88 \\ \hline 1.10 \end{array} \quad \begin{array}{r} 0.222 \\ +0.888 \\ \hline 1.110 \end{array} \quad \begin{array}{r} 0.2222 \\ +0.8888 \\ \hline 1.1110 \end{array} \quad \begin{array}{r} 0.22222 \\ +0.88888 \\ \hline 1.11110 \end{array} \quad \begin{array}{r} 0.222222 \\ +0.888888 \\ \hline 1.111110 \end{array}$$

The sum is clearly the infinite decimal 1.111111 . . . . in which all the digits are ones. To multiply these same decimals, we proceed by these steps:

$$\begin{array}{r} 0.2 \\ \times 0.8 \\ \hline 0.16 \end{array} \quad \begin{array}{r} 0.22 \\ \times 0.88 \\ \hline 176 \\ 176 \\ \hline 0.1936 \end{array} \quad \begin{array}{r} 0.222 \\ \times 0.888 \\ \hline 1776 \\ 1776 \\ 1776 \\ \hline 0.197136 \end{array} \quad \begin{array}{r} 0.2222 \\ \times 0.8888 \\ \hline 17776 \\ 17776 \\ 17776 \\ 17776 \\ \hline 0.19749136 \end{array}$$

We see that the product begins with 0.197, and more of the digits are fixed when more of the steps are taken.

At each step in the addition or multiplication just defined, we are using rational numbers. The rational numbers obey the five laws. As a result, the five laws are extended to the real numbers as well. For example, 0.2 + 0.8 and 0.8 + 0.2 lead to the same sum; 0.22 + 0.88 and 0.88 + 0.22 lead to the same sum; etc. So 0.222 . . . + 0.888 . . . and 0.888 . . . + 0.222 . . . lead to the same infinite decimal as sum. In other words, real numbers obey the commutative law of addition. The other four laws can be established by a similar argument.

We Still Have the Rational Numbers

We constructed the real number system in such a way that we have a number for every point on the number line. Among the points on the line are those to which we had previously assigned rational numbers. Let us refer to these as the rational points on the line. The real numbers assigned to the rational points form a subset of the real number system that is isomorphic to the rational number system. For all practical purposes, they are the "same" as the rational numbers. The symbol for a rational number is a fraction or a finite decimal. Such a symbol is easier to write and work with than an infinite decimal, so we use the rational number rather than the infinite decimal to represent the real number attached to a rational point. But first we must learn to recognize which of the real numbers belong to the rational points, and are therefore entitled to this simpler representation.

The infinite decimals that represent rational points are those which, after a finite number of digits in the decimal, simply repeat a fixed block of one or more digits over and over again. For example, the following decimals, in which the repeating block has been italicized, all represent rational numbers: 0.*3*333 . . ., 0.*12*1212 . . ., 2.37*45*4545 . . . . To prove this fact, we have to show, first, that every rational number, represented by a fraction, can be written as a repeating decimal. Then, vice versa, we must show that every repeating decimal can be written as a fraction.

To show that a fraction can be written as a repeating decimal, recall that we can convert a fraction into a decimal by long division, dividing the denominator into the numerator. In the long division process there is a step involving subtraction, after which we carry down a digit from the dividend. The dividend is an integer with a finite number of digits. So these are soon exhausted. Then we begin carrying down the zeros that appear after the decimal point. Consider what happens after we reach this stage of carrying down only zeros. In the subtraction step there is a remainder that is less than the divisor, and this remainder determines what the next number in the quotient will be. Since the remainders must be less than the divisor, the list of possible remainders is a restricted finite list. But, as we proceed with the division, we get an endless succession of remainders. So we cannot keep getting a different remainder each time. Sooner or later, a remainder that turned up before is repeated, and the division process begins to repeat itself.

To show that every repeating decimal can be written as a fraction, we shall work out one specific example showing how it is done. It will be clear that the method used can be employed with any repeating decimal whatever.

Suppose we find the fraction that represents the repeating decimal 2.7*15*1515 . . . . We first split this decimal into two parts, 2.7, and .0*15*1515 . . . ., separating the non-repeating part from the repeating part. The first part is the fraction $\frac{27}{10}$.

Now we find a fraction for the second part. First multiply it by ten, so that the repeating block will begin right after the decimal point. Let us call the result x, and remember that it is ten times as big as the number we are looking for; so after we find x we must divide by ten. $x = .151515 \dots$ Now multiply both sides of this equation by 100. This has the effect of moving the decimal point two places to the right.

We get $100x = 15.1515 \ldots$, which may also be written as $100x = 15 + .1515 \ldots$ But the decimal in this equation is none other than x all over again. So we may write $100x = 15 + x$. Taking x away from both sides, we find that $99x = 15$. Dividing by 99 on both sides, we find that $x = \frac{15}{99}$ or $\frac{5}{33}$. Now we divide by ten, to find that the second part of our original number is $\frac{5}{330}$. So the infinite decimal $2.7151515 \ldots$ is the sum of $\frac{27}{10}$ and $\frac{5}{330}$. Therefore $2.7151515 \ldots$ represents the rational number $\frac{896}{330}$ or $\frac{448}{165}$. To check the result, divide 165 into 448.

We Still Have a Field

The rational number system has the structure of a field. Enlarging the number system has not destroyed this structure, because the real number system, too, is a field. We can verify that it has the characteristics of a field, one by one. In the first place it is a commutative group with the operation of addition. The zero element in the group is the infinite decimal $0.0000 \ldots$, which we may write briefly as 0, and every infinite decimal has a negative, namely, the infinite decimal written with the same digits in the same order, but having the opposite sign attached. For example, the negative of $.333 \ldots$ is $-.333 \ldots$ It is also a ring, because the multiplication is distributive with respect to addition. The unity element may be written in two ways: $1.00000 \ldots$, or $0.9999 \ldots\ldots$ Moreover, every element except 0 has a reciprocal, so the system is a field. To find the reciprocal of an infinite decimal, we can use the successive finite decimals that approximate it, and divide each into the number 1. The quotients we get serve as successive approximations of the reciprocal, and, one by one, we can identify the digits in the decimal that represents it. In special cases we have simplified ways of doing it. For example, to find the reciprocal of $\sqrt{2}$, we write it first in fraction form as $\frac{1}{\sqrt{2}}$. The value is unchanged if we multiply by $\frac{\sqrt{2}}{\sqrt{2}}$, because this multiplier is equal to 1, or the unity element. But then we have $\frac{\sqrt{2}}{2}$, whose decimal equivalent is easily found by dividing 2 into the decimal for $\sqrt{2}$. The decimal for $\sqrt{2}$ begins as 1.414 . . . , so the decimal for its reciprocal begins as .707 . . . .

A Number in Every Nest

The system of real numbers has some special properties that the rational number system does not have. The most convenient way of expressing these properties is in terms of the picture we have set up of numbers as points on a line. When we represented the rational numbers as points on a line, we found that there were gaps that it left unfilled. That is, there were points on the line that had no numbers attached to them. The real number system was deliberately designed to eliminate this defect. In this system, not only do we have a point for every number. We also have a number for every point. There is a one-to-one correspondence between the real number system and the points on the number line. Because of this correspondence, we may think of the real numbers as the points on the line, and can describe the properties of the real number system in terms of relationships of the points on the line.

For example, by making every infinite decimal an element of the real number system, we assured the fact that there would be a real number for every such decimal. This property can also be described in terms of points on the line, as we have seen: an infinite decimal represents a nest of intervals on the line, and for every such nest, there is one and only one point that lies inside every interval of the nest. In this statement, the nest referred to is a nest associated with an infinite decimal. The successive intervals in such a nest have a special length, viz., 1, .1, .01, .001, etc., and their endpoints are always finite decimals. However, it is possible to form nests of intervals of a more general character by removing these restrictions on the lengths of the intervals and the locations of their endpoints. The only requirements for calling a sequence of intervals a nest are that the successive intervals lie one inside the other and that the lengths of the inner intervals shrink toward 0. It can be proved that in the real number system, all *nests* have the same property we have found for nests associated with infinite decimals: There is one and only one point that lies inside every interval of the nest. In this sense, every nest of intervals defines a single real number. Some of the other properties of the real number system that we shall now examine are closely related to this fact.

Infinite Series

Addition is an operation defined on a pair of numbers, so, initially, we can add only two numbers at a time. However, by performing one addition after another, we can extend the operation to include any finite number of numbers. For example, there is no difficulty about finding the sum of these numbers: 1 + 1 + 1 + 1 + 1 + 1. The sum is, of course, 6. However, if we permit the series of ones to go on indefinitely, in the infinite series 1 + 1 + 1 + 1 + 1 + 1 . . . ., then we run into trouble. The step by step addition that we can carry out with a finite number of terms doesn't work here, because it never comes to an end. We are left then with the question: Does it make any sense at all to try to add an infinite series of terms? The answer turns out to be that sometimes it does, and sometimes it doesn't. We can get clues to when an infinite series has meaning as a sum by re-examining an infinite series with which we are already familiar, namely, an infinite decimal.

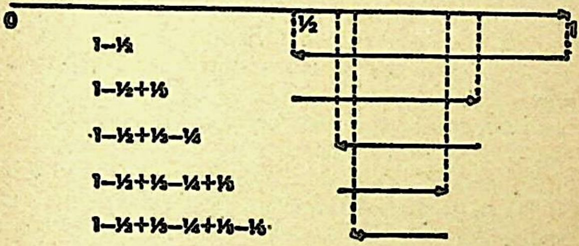
The infinite decimal .3333333 . . . . . . is really an infinite series in disguise. In fact, we may think of it as merely an abbreviated way of writing the infinite series .3 + .03 + .003 + .0003 + .00003 . . . . . . On page 97 we also interpreted the infinite decimal as a sequence of finite decimals, .3, .33, .333, .3333, ... These finite decimals are the sums we get when we add a finite number of terms in the infinite series, using the first term alone, then the first two terms, then the first three, and so on. We call these sums the partial sums of the series. We found that these partial sums come closer and closer in value to the fraction $\frac{1}{3}$, approaching this value as a limit, so we assigned the value $\frac{1}{3}$ to the infinite decimal. In a similar way, we can assign a meaning to some other examples of infinite series. *If the partial sums of a series come closer and closer, in the long run, to some definite number, approaching this number as a limit, then we can assign this limiting number as the sum of the infinite series.*

For example, let us consider the series, $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots$ in which each term is half the size of the term that it follows. The partial sums are, $1, 1\frac{1}{2}, 1\frac{3}{4}, 1\frac{7}{8}, 1\frac{15}{16}, 1\frac{31}{32}, \ldots\ldots$ In this case, the partial sums come closer and closer to the number 2. The first sum differs from 2 by 1. The second sum differs from 2 by $\frac{1}{2}$. The third sum differs from 2 by $\frac{1}{4}$. Successive partial sums differ from 2 by smaller and smaller amounts, and the difference can be made as small as we please if we add up enough terms in the series. So the partial sums approach the number 2 as a limit. We therefore can assign the number 2 as the sum of the series $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots.$

Let us try to use the same procedure with the series $1 + 1 + 1 + 1 + \ldots\ldots$ The partial sums for this series form the sequence, 1, 2, 3, 4, 5, 6, ... In this case, there is no number that the partial sums approach as a limit, because they keep increasing without limit. That is, we can get a partial sum to be larger than any number we wish by simply adding enough terms of the series. Because the partial sums do not approach a limit, we cannot assign any number as the sum of this series. So we see that not every infinite series has meaning as a sum. An infinite series has meaning as a sum only if the partial sums of the series approach a limit.

In that case we call the series a convergent series, and the sum of the series is the limit approached by the partial sums.

Now that we know that only some infinite series have meaning as a sum, the next logical question to ask is, "Which ones?" How do we recognize an infinite series that converges? We find that we can recognize some very easily by the fact that they define a nest of intervals. For example, consider the series $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \ldots$, in which the successive terms decrease toward 0 and the signs alternate between + and −. As we form the partial sums, let us locate them as points on the line of real numbers. The first partial sum is the number 1. Because the next term, $-\frac{1}{2}$, is negative, the second partial sum is to the left of the first. Adding the third term, $\frac{1}{3}$, brings us back to the right, but not all the way back, because $\frac{1}{3}$ is smaller than $\frac{1}{2}$. As we add in more and more terms, the point representing the partial sum oscillates back and forth, left and right, but never again as far to the left as before, and never again as far to the right. Now consider the intervals that have successive partial sums as their endpoints. The first interval is bounded on the right by 1 and on the left by $1 - \frac{1}{2}$. The second interval is bounded on the left by $1 - \frac{1}{2}$, and on the right by $1 - \frac{1}{2} + \frac{1}{3}$. The third interval is bounded on the right by $1 - \frac{1}{2} + \frac{1}{3}$, and on the left by $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4}$. Each new interval is inside the one that it follows. Moreover, the width of the intervals is shrinking toward 0. The intervals form a nest, and, as we know, there is one and only one point inside the nest. The partial sums are crowding in toward this point as a limit,



so the series $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots$ is a convergent series.

A similar argument applies to any series in which the terms decrease toward zero, and the signs are alternately + and −. Every such series defines a nest of intervals, and converges to the single point that is inside the nest. This result, which we have established for the real number system, does not hold for the rational number system, because there it is not true that every nest of intervals has a point inside.

Another type of series that is easy to analyze is one in which all terms are positive, and the partial sums are bounded, that is, they are always less than some fixed number. The series $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots$ is an example of this type, because all partial sums are less than 2. Let us represent the general series of this type by $a_1 + a_2 + a_3 + \ldots$. where the subscripts 1, 2, 3, etc. are labels to identify the position of each term in the series. The partial sums are $a_1, a_1 + a_2, a_1 + a_2 + a_3, \ldots$ Since we get each partial sum from the one that precedes it by adding a positive number, the partial sums form an increasing sequence. Let us represent the partial sums by the symbols $S_1, S_2, S_3, S_4, \ldots$. That is, $S_1 = a_1$, $S_2 = a_1 + a_2$, $S_3 = a_1 + a_2 + a_3$, and so on.

If we represent the partial sums as points on the line of real numbers, since the numbers are increasing we get a succession of points *moving gradually to the right.* S_2 is to the right of S_1, S_3 is to the right of S_2, and so on. However, since the sums are all less than some fixed number, they cannot move too far to the right. If the fixed number is represented by K on the number line, its position is an upper boundary beyond which the sequence of partial sums cannot go. Now we show how we can use K and the partial sums to define a nest of intervals. The points S_1 and K are the ends of an interval. This is the first interval of the nest. All the S points after S_1 lie to the right of S_1 and to the left of K, so they are inside this interval. Let us divide this interval in half. The points in the sequence of S's are moving to the right. Either they finally enter the second half of the interval, or they do not. If they do enter the second half, once they enter they stay there, because they keep moving to the right, and never get past K. In that case we use this second half of the interval as the second interval of the nest. If the S points never enter the second half, that means they always remain in the first half. Then we use the first half as the second interval of the nest. Now we repeat the process. We divide the second interval of the nest in half, and pick one of the halves as the third interval of the nest. We choose the second half if the S's ultimately enter it. We choose the first half if they never enter the second half.

In this way we get a sequence of intervals with the characteristics of a nest: the intervals are one inside the other, and the width of the intervals is shrinking toward zero. Then there is a single real number that lies inside all the intervals of the nest. Since we know from the way in which we chose the intervals of this nest that the partial sums enter and remain in each of them, then the partial sums converge toward this single number as a limit. Therefore an infinite series of positive terms whose partial sums are bounded converges to a limit. This result can also be restated in terms of the sequence of partial sums alone, without reference to the series from which the sums were derived: Every increasing sequence of numbers that is bounded on the right converges to a limit.

The two types of convergent series just examined are only special cases. It is not difficult, however, to find a criterion by which all convergent series can be recognized. Suppose the series is represented by $a_1 + a_2 + a_3 \ldots$, where the terms may be either positive or negative. Whenever we form a partial sum, we are breaking the series up into two parts. The first part consists of a finite number of terms at the beginning of the series, taken in order, and added to get the partial sum. Let us call this part the head end of the series. The second part consists of all the remaining terms, that are not used to form this partial sum. Let us call this part the tail end of the series. We can designate the successive partial sums, formed from the head end as more and more terms are included in it, by S_1, S_2, S_3, etc. Let us call the corresponding tail ends T_1, T_2, T_3, etc. Then

$S_1 = a_1$, with tail end T_1: $a_2 + a_3 + a_4 \ldots$

$S_2 = a_1 + a_2$, with tail end T_2: $a_3 + a_4 + a_5 + \ldots$

$S_3 = a_1 + a_2 + a_3$, with tail end T_3: $a_4 + a_5 + a_6 + \ldots$

As we move along the sequence of partial sums, S_1, S_2, etc., we take one term at a time from the tail end and transfer it to the head end to be included in the partial sum. As a result, as n increases, the value of the partial sum, S_n, keeps changing. If the last term that is transferred is positive the change is an increase. If the term is negative, the change is a decrease. If we disregard the sign of the term we get a positive number that tells us the size of the change without regard to whether it is a decrease or increase. This positive number is called the absolute value of the change. Thus, an increase by $\frac{1}{2}$ or a decrease by $\frac{1}{2}$ both have an absolute value of $\frac{1}{2}$.

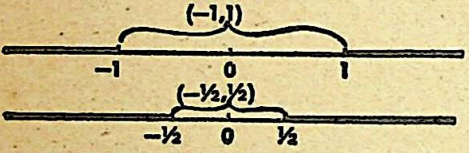
The series converges if the partial sums approach a limit. If the partial sums approach a limit, they become more and more nearly equal to that limit. This means that, as more and more terms are included in the partial sum, the sum changes less and less. And if n is taken large enough, the partial sum S_n is so close to the limit, that it changes very little no matter how many more terms are transferred from the tail end of the series to the head end. In fact, if n is large enough we can keep the absolute value of this change in the partial sum as small as we please. The partial sum S_n is equal to $a_1 + a_2 + \ldots + a_n$. The tail end T_n consists of the series $a_{n+1} + a_{n+2} + a_{n+3} + \ldots$. If we transfer the first p terms from the tail end to the head end, we add to S_n the sum $a_{n+1} + a_{n+2} + \ldots + a_{n+p}$. So, if the series converges, the absolute value of this sum must shrink toward zero as n increases, no matter how many terms from the tail end it includes. The converse is also true. If the absolute value of the sum of the first p terms of the tail end of the series shrinks to zero as n increases, no matter how large p is, then the series converges. This fact can be proved by showing that, under these conditions, there is a nest of intervals into which the partial sums crowd, so that they converge on the single point that is inside the nest. This criterion for a convergent series is known as the Cauchy criterion, and may be summed up somewhat carelessly in these words: in the real number system, an infinite series converges to a limit if and only if it has a shrinking tail end.

Limit Points and Neighborhoods

So far we have passed through four stages in the enlargement of our number system. We started with the system of natural numbers. Then by successive extensions, we obtained the integers, the rational numbers, and the real numbers. This sequence of extensions, besides showing us how our notion of number has been evolving, also served to introduce us to a variety of mathematical structures. In the natural number system, we first encountered the structure which we have labeled a "*number system*" and which is distinguished by its obedience to the five laws. In the system of integers we found an example of both a *group* and a *ring*. The system of rational numbers was our first example of a *field*. Now, in the system of real numbers, we shall get acquainted for the first time with another type of structure, a *topological space*. A real number is an infinite decimal, and an infinite decimal is an infinite sequence of finite decimals. This fact compelled us to look into the question of when an infinite sequence converges to a limit. Now, as we examine more closely the notion of a limit, it will lead us to the concept of a topological space.

Suppose we look at the sequence of numbers, $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$, in which the nth number is $\frac{1}{n}$. As n increases, these numbers become smaller and smaller and approach 0 as a limit. If we represent the numbers in the sequence as points on a line, the fact that they approach 0 as a limit shows up in the fact that the points crowd in toward the zero point. This crowding can be described without any reference to the order in which the numbers are arranged in the sequence. In fact, let us discard the sequence and merely think of the numbers as the set of points on the line. Let us give this particular set the name A. To define in what sense the points in the set A crowd in toward the zero point, we first introduce the notion of a neighborhood of a point.

A neighborhood of a point consists of all the points that surround it and are close to it. To make this concept more precise, we have to indicate what we mean by *close*. So we specify degrees of closeness, and each such specification defines a particular neighborhood. For example, we define a neighborhood of zero when we say it consists of all the points whose distance from zero is less than 1. This neighborhood includes all points that are larger than (to the right of) -1 but less than (to the left of) $+1$. In the diagram, we see this neighborhood as the interval between -1 and $+1$, not including its endpoints. We represent this neighborhood by the symbol $(-1, 1)$, in which its boundary points are shown. We define another neighborhood of zero when we specify that it consists of all points whose distance from zero is less than $\frac{1}{2}$. This is a smaller neighborhood than the first one, and is included within it. All of its points lie between $-\frac{1}{2}$ and $+\frac{1}{2}$. We can represent it by the symbol $(-\frac{1}{2}, \frac{1}{2})$.



Neighborhoods of zero

By picking any positive distance d, we can define a neighborhood of zero as the set of points lying between $-d$ and $+d$, and we represent it by the symbol $(-d, d)$. So we see that the zero point is surrounded by a multitude of neighborhoods of many sizes.

Now we can say exactly what is meant by the fact that the points in the set A $\left(A = \left\{1, \frac{1}{2}, \frac{1}{3}, \ldots \frac{1}{n}, \ldots\right\}\right)$ crowd in toward zero: If we pick any neighborhood of zero, no matter how small, all but a finite number of these points lie within that neighborhood. If we choose smaller and smaller neighborhoods, we can exclude some of the points of the set A from the neighborhood. But, no matter how many we exclude, all but a finite number of the points are still inside. This is the meaning of the fact that 0 is the limit point of the set A.

Now let us examine a set in which we find crowding of a different kind. Let B stand for the set

$$\left\{1+\tfrac{1}{2}, 1+\tfrac{1}{3}, 2-\tfrac{1}{3}, 1+\tfrac{1}{4}, 2-\tfrac{1}{4}, \ldots, 1+\frac{1}{n}, 2-\frac{1}{n}, \ldots\right\}.$$

In this set, we find the points crowding around two points of the line, 1, and 2. Here we cannot say that every neighborhood of the point 1 includes all but a finite number of points in the set. In fact, while the neighborhood $(\frac{1}{2}, 1\frac{1}{2})$ includes an infinite number of points of the set B, it also excludes an infinite number, because there are multitudes of them clustering around the point 2. To distinguish this case from the preceding case, we call the points 1 and 2 *cluster points* of the set B. A point is a cluster point of a set of points if every neighborhood of the point includes an infinite number of points of the set. As we see, a set may have more than one cluster point. If a set is confined within a finite interval, and has only one cluster point, then that single cluster point is the limit point of the set.

As a contrast to the behavior of a limit point or a cluster point, let us examine the point -1 in relation to the set A. Surround the point -1 by the neighborhood that extends from $-1\frac{1}{2}$ to $-\frac{1}{2}$. Within this neighborhood, represented by the symbol $(-1\frac{1}{2}, -\frac{1}{2})$, there are no points of the set A. That is, the neighborhood $(-1\frac{1}{2}, -\frac{1}{2})$ excludes all points of the set A. On the other hand, there is no neighborhood of the point 0 that excludes all points of the set A. Similarly, there is no neighborhood of the point 1 that excludes all points of the set B, and there is no neighborhood of the point 2 that excludes all points of the set B. Because they have this property, a limit point and a cluster point are both examples of what we call a point of adherence of a set.

A point is called a *point of adherence* of a set of points if no neighborhood of the point excludes all points of the set. A point need not be a limit point or a cluster point of a set in order to be a point of adherence of the set. For example, the point $1\frac{1}{2}$ is not a cluster point of the set B, but it is a point of adherence. No neighborhood of $1\frac{1}{2}$ can exclude all points of the set B because the point $1\frac{1}{2}$ is itself a member of the set. The points of adherence of a set include all members of the set as well as cluster points that are not in the set. They are the points that cling to a set by either being in it or being crowded by it.

Closed Sets and Open Sets

With the help of the concept of point of adherence, we now define two special kinds of sets of points on the line. A set that contains all of its points of adherence is called a *closed set*. The set A, discussed in the paragraphs above, is not a closed set, because the point 0 is a point of adherence of the set but does not belong to it. However, if we enlarge the set by including 0 as a member, then the enlarged set is closed. Similarly, the set B is not closed. But the enlarged set formed by uniting B with the set $\{1, 2\}$ is closed. Another example of a closed set is the set of points between 0 and 1, *including the points 0 and 1*. Such a set is called a *closed interval*. The set which contains *all* the points on the line of real numbers is also a closed set. We can count the empty set as a closed set, too. It certainly includes all its points of adherence, because there aren't any.

If we delete from the real number line all the members of some closed set, what is left is called an *open set*. Using the terminology defined in Chapter II, we can say that an open set is the complement of a closed set. To see what an open set is like, we have to think of it in relation to the closed set which is its complement. Suppose S is an open set, and C is the closed set which is its complement. Any point of the open set S is not in the closed set C. Therefore any point in the open set S is not a point of adherence of C. (If it were, it would have to be in C, by the definition of a closed set.)

But if a point is not a point of adherence of *C*, then some neighborhood of the point contains no points of *C*. This neighborhood, which contains no points of *C*, must then be part of the open set *S*, which consists of all the points that are not in *C*. So we have discovered that every point of an open set is surrounded by an entire neighborhood that is also in the open set. This is a distinguishing feature of open sets on a line.

A neighborhood is itself an example of an open set. For example, the set of points between 0 and 1, *not including 0 and 1*, is an open set. It is called an *open interval* to distinguish it from the closed interval which does include both end points. Another example of an open set is the set formed by uniting into one set all points in any collection of open intervals. The whole line of real numbers is also an open set, because it is the complement of the empty set, which is a closed set. The empty set, too, is an open set, because it is the complement of the whole line, which is a closed set. We see then that the empty set and the whole line are both open and closed.

There are some sets that are neither open nor closed. For example, the set of points between 0 and 1, including the endpoint 0, but not including the endpoint 1, is neither open nor closed. It is not closed, because 1 is a point of adherence of the set, but doesn't belong to it. It is not open, because 0 is a point of adherence of the complement of the set, but does not belong to the complement, so the complement is not closed.

The collection of all open sets on the line has the following properties, some of which we have already noted:

1) The whole line, as well as the empty set, are open sets.
2) The union of any number of open sets is also an open set.
3) The set of points common to two open sets (the intersection of the two sets) is an open set.

Neighbors Make a Neighborhood

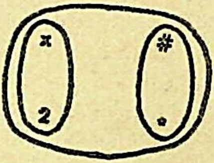
The significance of the concept of an open set is that it permits a generalization of the concept of neighborhood.

We have already seen that a neighborhood is an open set. Let us agree to extend the word neighborhood to include every open set. The effect of this generalization is to separate the notion of neighborhood from the idea of distance. Then a neighborhood becomes simply a collection of neighbors, with the characteristics of an open set. From this point of view, the whole line may be viewed as a system of interlocking neighborhoods or open sets. The interlocking neighborhoods on the line determine what is called its topological structure, and make it a topological space. They fix a pattern of relationships within the space just as the interlocking stitches in a sweater fix the pattern of the sweater.

In general, any set of objects is called a topological space if a collection of its subsets are singled out so that the collection has the three properties we found in the open sets on the line: 1) The whole space and the empty set belong to the collection; 2) The union of any number of sets in the collection is also in the collection; 3) The intersection of any two sets in the collection is also in the collection. When these three conditions are satisfied, the sets in the collection are called the "open sets" of the "space."

Under this definition, any collection of objects can be converted into a topological space, usually in more than one way. For example, let us consider the set $\{x, 2, \#, *\}$. Its elements were chosen arbitrarily, so they have no relationship to each other beyond the fact that they happen to have been thrown together in the same set. The set is a loose aggregation, and has no structure. However, the set acquires a structure, and the elements become related to each other as neighbors, as soon as we single out certain subsets that will be called neighborhoods or open sets.

For example, we might specify that these four sets should constitute the collection of open sets: $\{x, 2, \#, *\}$, $\{x, 2\}$, $\{\#, *\}$, and $\{\ \}$. This collection satisfies the three requirements listed above. 1) The whole space, $\{x, 2, \#, *\}$, and the empty set, $\{\ \}$, are members of the collection. 2) The union of any number of sets in the collection is also in the collection. For example, the union of $\{x, 2\}$ and $\{\#, *\}$ is $\{x, 2, \#, *\}$, which is in the collection. 3) The intersection of any two sets in the collection is also in the collection. For example, the intersection of $\{x, 2\}$ and $\{\#, *\}$ is the empty set $\{\ \}$, which is in the collection. Since the requirements are met, this collection of "open sets" defines a topological structure for the set $\{x, 2, \#, *\}$ and converts it into a topological space. The way the topological structure relates the elements of the space to each other as members of neighborhoods is indicated in the diagram below, where each open set is represented by a loop enclosing its members.

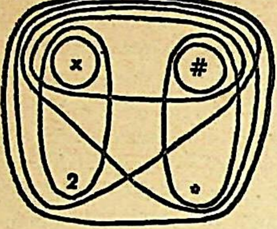


We can give the same set another, different topological structure by picking out other subsets to use as "open sets." We might, for example, decide that we want the collection of open sets to include all those listed above, and in addition the set $\{x, \#\}$. The inclusion of this one additional set among the open sets compels us to include more sets, in order to meet the three requirements. The union of $\{x, \#\}$ and $\{x, 2\}$ is $\{x, 2, \#\}$. To meet requirement 2), we must include it among the open sets. The intersection of $\{x, \#\}$ and $\{x, 2\}$ is the set $\{x\}$. To meet requirement 3), we must classify it as an open set. For similar reasons, we have to include $\{x, \#, *\}$ and $\{\#\}$ among the open sets. We find that all three requirements are met now by the enlarged collection consisting of these nine sets:

$$\{x, 2, \#, *\},\ \{x, 2\},\ \{\#, *\},\ \{\ \},$$
$$\{x, \#\},\ \{x, 2, \#\},\ \{x, \#, *\},\ \{x\},\ \{\#\}.$$

The enlarged collection of open sets gives the space a more complicated structure of interlocking neighborhoods, as shown in the diagram on page 122.

A third topological structure can be defined for the same



set by specifying that *every* subset shall belong to the collection of open sets. This definition meets the three requirements, because the whole set and the empty set are both subsets; the union of any number of subsets is a subset; and the intersection of any two subsets is a subset. Under this definition, the set of four elements becomes a topological space containing sixteen open sets, including the empty set. The three topological structures we have defined convert the same set of four elements into three different topological spaces. They are different as spaces because, in each of the topological structures, the elements hang together differently as members of interlocking neighborhoods.

Here is another example of a topological space. Consider all the straight lines that can be drawn in a plane parallel to some fixed direction. We can convert this set of lines into a topological space, in which each of the lines is a "point" of the space, by defining for the space a structure of interlocking open sets. Let an open set consist of a set of lines between any two lines, or a union of any number of such sets. In this space, a set of lines between any two lines is analogous to what we called an open interval in the real number system, namely, a set of points between any two points.

A more familiar looking example of a topological space is the circumference of a circle, in which arcs play the same role that intervals do on a straight line. On a circle, there are two arcs joining any two of its points. The interior (excluding the endpoints) of the arc may serve as an "open" arc just as the interior of a segment serves as an open interval on a straight line. An open set on the circle can then be defined as the union of any number of such open arcs. With this definition, the circle is furnished with a topological structure that meets the three requirements listed on page 120. This topological structure on a circle can also be derived from the topological structure on a straight line in which each open set is a union of open intervals. Take a segment of a straight line, and loop it to make a circle by joining the two ends together. Then open intervals on the straight line become open arcs on the circle.

By isolating the topological structure of the real number system for separate study, mathematicians have obtained a deeper insight into the characteristics of that system. They could separate from each other those properties that the system has by virtue of the fact that it is a topological space, those that it has by virtue of the fact that it is a field, and those that it has because it happens to be both. The separate study of topological structures also develops a body of knowledge that is applicable to other topological spaces, no matter what the elements of these spaces may be.

Rubber-Sheet Geometry

Now that we have introduced the notion of a topological space, it is relevant to ask, "When are two topological spaces essentially the same?" The problem is analogous to the problem we encountered when we were talking about groups, rings, or fields, and has a similar answer. We said that two fields are essentially the same, or are isomorphic, if there is a one-to-one correspondence between them that preserves the field structure embodied in the operations of addition and multiplication. Similarly, we say that two topological spaces are essentially the same, or are *homeomorphic*, if there is a one-to-one correspondence between them that preserves the topological structure embodied in the system of interlocking open sets. That is, two topological spaces are homeomorphic if, under a reversible mapping which establishes a one-to-one correspondence between them, an open set in either space has as its image an open set in the other. For example, imagine the circumference of a circle being stretched like a rubber band, into a distorted shape. Each point of the circle may take up a new position. Distances between points change, but the interlocking system of open sets remains. As a result the distorted curve obtained is homeomorphic to the original circle. Because the topological structure of a space is not changed when it is stretched without tearing, the study of topological structure has sometimes been referred to as "rubber-sheet geometry."

DO IT YOURSELF

1. By an argument similar to the one used on page 94, prove that there is no rational number equal to the square root of 3.
2. Find the repeating decimal that represents the rational number $\frac{2}{13}$.
3. What rational number is represented by the repeating decimal .181818. . . . ?
4. Define a topological structure for the set of five elements $\{a, e, i, o, u\}$ by designating some but not all of the subsets as "open" sets. (See example on page 120.) Be sure your open sets meet the requirements for a topological structure listed on page 120. Which sets are closed sets (complements of open sets) in this structure?
5. Every even integer has the form $2k$, and every odd integer has the form $2k + 1$, where k is an integer. Use this notation to prove that the square of every even integer is an even integer, and the square of every odd integer is an odd integer.
6. Find the repeating decimal that represents

   a) $\frac{1}{7}$ b) $\frac{3}{7}$ c) $\frac{5}{11}$ d) $\frac{2}{17}$ e) $\frac{5}{12}$
7. Find the rational number represented by the repeating decimal:

   a) .555. . . b) .777. . . c) .060606. . .
   d) .050505. . . e) .121212. . . f) .161616. . .
   g) .123123123. . . h) .025025025. . . i) 6.2575757. . .
   j) 3.2777. . .

CHAPTER VI

Spilling Into the Plane

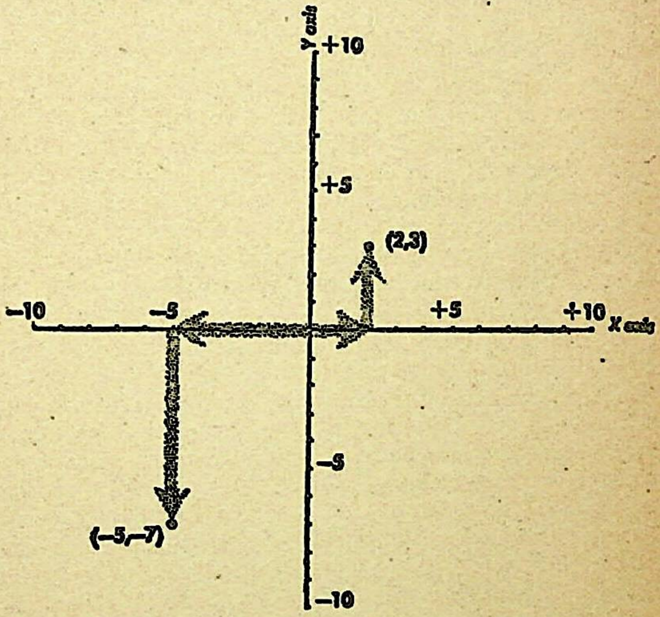
IN THE system of rational numbers we were able to solve the equation $x^2 - 1 = 0$, but we were unable to solve the equation $x^2 - 2 = 0$. We remedied this defect by constructing a larger number system, the system of real numbers, in which the equation $x^2 - 2 = 0$ does have a solution. In fact, it has two solutions in the real number system, viz., $+\sqrt{2}$, and $-\sqrt{2}$. But the real number system has some defects of its own. For example, although the equation $x^2 + 1 = 0$ does not look any more complicated than the other two equations mentioned above, it has no solutions in the real number system. We can see why this is so, by restating the question that the equation asks us. First we add -1 to both sides of the equation, and we get the equivalent equation $x^2 = -1$. This equation says, "Find a number which, when multiplied by itself, gives -1 as the product." The real number 0 does not meet the requirement, because 0 multiplied by itself gives 0 as the product. A positive real number can not meet the requirement, because a positive number multiplied by itself gives a positive number as the product. A negative real number cannot answer the question either, because a negative number multiplied by itself also gives a positive product. For example $(-1) \cdot (-1) = +1$. So there is no real number that can satisfy the equation $x^2 + 1 = 0$, or $x^2 = -1$. Our next goal is to construct an expanded number system in which this equation does have a solution.

At each stage in the expansion of the number system so far we have represented numbers pictorially as points on a line. With the construction of the real number system, we finally achieved our purpose of having a number for every point on the line. Another possible goal for the next step in the expansion of the number system would be to spill over into the plane, by constructing a number system that gives us a number for every point in the plane. It turns out that these two goals coincide. We shall reach the goal of finding a solution to the equation $x^2+1=0$ by building a number system that supplies a number for every point in the plane.

Number Pairs and Arrows

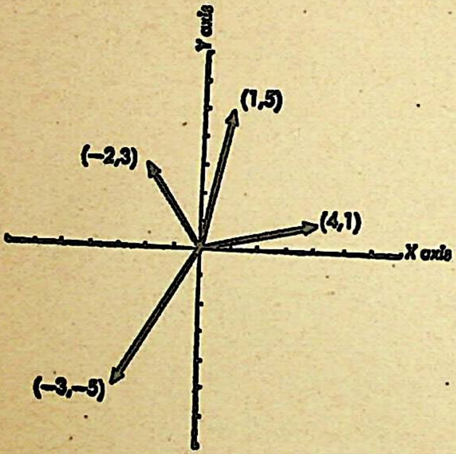
The construction of a number system that has a number for every point in the plane will be carried out in two installments. We shall do only half of the job in this chapter, by creating an appropriate system of elements, and defining an addition operation on these elements. We shall finish the job in Chapter VIII when we define a multiplication operation for these elements. The elements we shall use are ordered pairs of real numbers, like $(0, 1)$, $(-2, 3\frac{1}{2})$, or $(1, \sqrt{2})$. We assign such a pair to each point in a plane by the familiar method used in elementary algebra to identify the co-ordinates of a point on a graph. First we draw in the plane a straight line called the x axis. We furnish it with a scale on which measurements can be made by assigning a real number to each point on the line, in the manner described in the preceding chapters. Then we draw another line passing through the 0 of the x axis, and crossing the axis at right angles. This new line is called the y axis. The point where the axes cross is called the *origin*. We put a scale on the y axis, too, using the origin as its zero point, and putting the positive numbers on the upper half of the axis.

Now we assign a pair of numbers to each point in the plane in the following way. Drop a perpendicular from a given point to the x axis. We can reach the given point from the origin by moving toward it in two steps. First move from the origin to the foot of the perpendicular. The distance moved supplies the first number of the ordered pair. It is a positive number if the motion is to the right of the origin. It is a negative number if the motion is to the left of the origin. Now, from the foot of the perpendicular, move along the perpendicular to the given point. The distance moved in this second step supplies the second number of the ordered pair. It is a positive number if the motion is up from the x axis. It is a negative number if the motion is down from the x axis. The pair of numbers is written inside parentheses. The first number in the pair is called its x component. The second number is called its y component. The ordered pairs that belong to some points in the plane are shown in the diagram below. Notice that the ordered pair (0, 0) belongs to the origin. For all points on the x axis, the y component is zero. For all points on the y axis, the x component is zero. We are going to convert



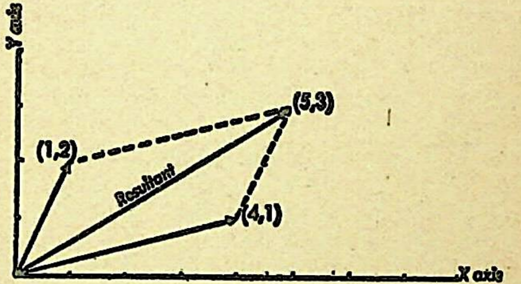
the system of ordered pairs of real numbers into a number system in its own right by giving it the necessary structure.

The point to which an ordered pair is assigned may be thought of as a picture of the pair. We can also associate another kind of picture with each pair in the following way. Draw an arrow from the origin to the point, with the arrowhead pointing away from the origin. We may think of the arrow as the second picture of the ordered pair. Each ordered pair, except (0, 0) has an arrow of its own. The arrow has a definite length, and a definite direction. We can even assign an arrow to the pair (0, 0) by giving it an arrow whose length is zero. However, we cannot assign any direction to this zero arrow.



Addition of Arrows

The representation of ordered pairs as arrows is very helpful, because it guides us to an appropriate definition of addition for our system of ordered pairs. There are many practical situations where we encounter just such arrows, and where there is a natural kind of addition that takes place. For example, in physics, a force can be represented as an arrow. The length of the arrow indicates the strength of the force, and the direction of the arrow indicates the direction of the force. If two forces act on a body at the same point, the effect is the same as though the body were acted on by a single force called the *sum* or *resultant* of the two forces. One way to find this sum or resultant is to draw the two forces as arrows at the origin, and then complete the parallelogram that has the two arrows as sides. The diagonal of the parallelogram that can be drawn from the origin is the resultant of the two forces.



However, there is another, simpler way of getting the resultant, too. Each force, considered as an arrow, has an x component and a y component. To add the two forces, all we have to do is add their x components separately, and add their y components separately. For example, the forces being added in the diagram above can be represented as arrows, or as the ordered pairs, (4, 1), and (1, 2). The x components are 4 and 1, and the y components are 1 and 2. Adding them separately, we find that the resultant or sum of the forces belongs to the ordered pair (5, 3). This example suggests how we should define addition of ordered pairs, if we want them to be useful for the solution of practical problems. We therefore give the following definition of addition in the new system: To add two ordered pairs, add their components separately. In symbols, this definition says, $(a, b) + (c, d) = (a + c, b + d)$. Because the sum of two ordered pairs is also an ordered pair, this addition is a binary operation.

There are several facts that we can observe about this binary operation immediately. First, it is associative. That is, $[(a, b) + (c, d)] + (e, f) = (a, b) + [(c, d) + (e, f)]$. This follows from the fact that we carry out the addition by adding the components, which are real numbers, and addition of real numbers is associative. For example,

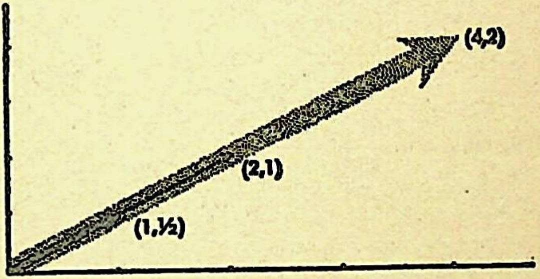
$$[(2, 3) + (3, 1)] + (4, 8) = (5, 4) + (4, 8) = (9, 12).$$

But $(2, 3) + [(3, 1) + (4, 8)] = (2, 3) + (7, 9) = (9, 12)$, too. Secondly, the addition of ordered pairs is commutative. That is $(a, b) + (c, d) = (c, d) + (a, b)$. This follows from the fact that addition of real numbers is commutative. For example, $(2, 3) + (3, 1)$ and $(3, 1) + (2, 3)$ both yield the same sum, $(5, 4)$. Third, the system has a zero element, because $(0, 0) + (a, b) = (0 + a, 0 + b) = (a, b)$. Finally, for each element in the system, there is a negative in the system, too, with the usual property that a negative is supposed to have. That is, the sum of any ordered pair and its negative is equal to the zero element. The negative of any ordered pair (a, b) is the ordered pair $(-a, -b)$. For example, the negative of $(2, 3)$ is $(-2, -3)$ because $(2, 3) + (-2, -3) = (0, 0)$. With these four characteristics, the system of number pairs we have constructed meets all the requirements for being an abelian group, with the group operation denoted by the plus sign.

With this observation, we complete the first half of the job of converting these elements into a number system. We shall go on with the second half of the job in Chapter VIII. Meanwhile we take a detour along a road pointed out by our system of arrows. As we follow this road we shall encounter some more mathematical structures that play an important part in modern mathematics.

Stretching the Arrows

In the diagram below we see an arrow representing the ordered pair (2, 1). If the arrow is stretched so that its length is doubled while its direction remains unchanged, we get a new arrow (the shaded arrow in the diagram). If the arrow is contracted to half its original length we also get an arrow pointing in the same direction. In the first case, the length of the arrow was multiplied by a factor of 2. In the second case it was multiplied by a factor of $\frac{1}{2}$. Because of this fact, it is natural to think of the operation of stretching or shrinking an arrow as a kind of multiplication of the arrow by a real number. To see how we may define the operation in terms of ordered pairs, notice that when the arrow (2, 1) was doubled, each of its components was doubled, giving the arrow (4, 2) as the result. So we define a special kind of multiplication for our system of ordered pairs as follows. To multiply an ordered pair of real numbers by a real number, multiply each of its components separately by that real number. In symbols, the definition says $a \cdot (b, c) = (a \cdot b, a \cdot c)$. It is important to notice that this kind of multiplication differs in one important respect from all the other multiplications defined so far in this book. In each of the earlier cases, the multiplication was an operation on two elements drawn from the same system. In this case, the two elements being multiplied are drawn from *two different systems*. One of the



multipliers is drawn from the field of real numbers. The other multiplier is drawn from the abelian group consisting of *ordered pairs* of real numbers. In order to sharpen the distinction between these two systems, we introduce some special names for them. We call the elements of the real number field *scalars*. The ordered pairs, or their pictorial representations as points in the plane or as arrows, will be called *vectors*. The kind of multiplication in which a vector is multiplied by a scalar to form a product which is also a vector is called *scalar multiplication*. We can verify from our definitions that scalar multiplication is *distributive with respect to vector addition*. That is, if s is a scalar, and (a, b) and (c, d) are vectors, then $s \cdot [(a, b) + (c, d)] = s \cdot (a, b) + s \cdot (c, d)$. To prove this rule, observe first that the expression on the left instructs us to do the vector addition first, and the scalar multiplication afterwards. So $s \cdot [(a, b) + (c, d)] = s \cdot (a + c, b + d) = (s \cdot [a + c], s \cdot [b + d]) = (s \cdot a + s \cdot c, s \cdot b + s \cdot d)$. The expression on the right instructs us to do two scalar multiplications first, and then add the results by vector addition. Following these instructions, we find that $s \cdot (a, b) + s \cdot (c, d) = (s \cdot a, s \cdot b) + (s \cdot c, s \cdot d) = (s \cdot a + s \cdot c, s \cdot b + s \cdot d)$. The proof is completed by observing that the two results are the same.

Scalar multiplication also satisfies another distributive law. It is distributive with respect to the addition of scalars. That is, if s and t are scalars, and (a, b) is a vector, then $(s + t) \cdot (a, b) = s \cdot (a, b) + t \cdot (a, b)$. To prove this rule, we observe that carrying out the indicated operations on both sides of the equals sign leads to the same result: $(s + t) \cdot (a, b) = ([s + t] \cdot a, [s + t] \cdot b) = (s \cdot a + t \cdot a, s \cdot b + t \cdot b)$. $s \cdot (a, b) + t \cdot (a, b) = (s \cdot a, s \cdot b) + (t \cdot a, t \cdot b) = (s \cdot a + t \cdot a, s \cdot b + t \cdot b)$.

Scalar multiplication also obeys a *mixed associative law*, in which two types of multiplication appear: 1) scalar multiplication, in which a scalar multiplies a vector, and 2) multiplication of scalars, in which a scalar multiplies a scalar. This law says that, if s and t are scalars, and (a, b)

is a vector, then $(s \cdot t) \cdot (a, b) = s \cdot [t \cdot (a, b)]$. To prove this law, observe that $(s \cdot t) \cdot (a, b) = ([s \cdot t] \cdot a, [s \cdot t] \cdot b) = (s \cdot t \cdot a, s \cdot t \cdot b)$. But $s \cdot [t \cdot (a, b)] = s \cdot (t \cdot a, t \cdot b) = (s \cdot [t \cdot a], s \cdot [t \cdot b]) = (s \cdot t \cdot a, s \cdot t \cdot b)$.

Scalar multiplication also has the property that the number 1, which is the unity element for multiplication of a scalar times a scalar, is also the unity element for multiplication of a scalar times a vector. This is seen from the fact that, if (a, b) is a vector, $1 \cdot (a, b) = (1 \cdot a, 1 \cdot b) = (a, b)$.

We shall now summarize these properties of scalar multiplication in a new abbreviated notation. So far we have always written a vector as an ordered pair, in which its two components are put on display. In the abbreviated notation, we represent a vector by a single symbol, with a little arrow over it to remind us that it stands for a vector. Thus, $\vec{x}$, $\vec{y}$, and $\vec{z}$ represent vectors. Symbols like r, s, and t, written without arrows, will represent scalars. In this notation, the properties of scalar multiplication are expressed in this form:

I. Distributive laws: $r \cdot (\vec{x} + \vec{y}) = r \cdot \vec{x} + r \cdot \vec{y}$

$(r + s) \cdot \vec{x} = r \cdot \vec{x} + s \cdot \vec{x}$

II. Mixed Associative law: $r \cdot (s \cdot \vec{x}) = (r \cdot s) \cdot \vec{x}$

III. Unity element: $1 \cdot \vec{x} = \vec{x}$

Because of the characteristics we have observed in the system of ordered pairs (pictured as arrows), this system serves as an example of the kind of mathematical structure that is called *a vector space*. The name vector space is used to describe any system of elements that has these properties: 1. It is an abelian group. 2. There is associated with this group another system of elements that is a field, and is called the scalar field. A scalar multiplication exists, in which a scalar from the field multiplies a vector from the vector space, and the product is a vector in the vector space. 3. The scalar multiplication has properties I, II, and III listed above.

Some Other Vector Spaces

Vector spaces won recognition as a special kind of structure worthy of separate study when mathematicians realized that there are many systems that have this kind of structure. In the example that we have been observing, each element in the vector space was an ordered pair of real numbers. We can construct another vector space, by using as elements ordered triples instead, like (1, 3, −2). In this system, each element has three components. We could define vector addition and scalar multiplication for this system in the same way that we did for the system of ordered pairs: To add two vectors, add their separate components separately; to multiply by a scalar, multiply each component separately by that scalar. For example, $(2, 3, -5) + (1, 4, 2) = (3, 7, -3)$. And $2 \cdot (2, 3, -5) = (4, 6, -10)$. With these definitions of vector addition and scalar multiplication, the system of triples meets all the requirements for being a vector space. The system of ordered pairs is an example of a vector space of two dimensions, and is represented pictorially by points in a plane. The system of ordered triples is an example of a vector space of three dimensions, and can be represented pictorially by the points in three dimensional space. By using ordered quadruples as elements, we can construct a vector space of four dimensions. In general, we can construct a vector space of n dimensions by using as elements ordered sets of real numbers with n numbers in each set. If we take $n = 1$, the vectors are single real numbers, and the vector space is one dimensional. In other words, the real number system may be thought of as a vector space of one dimension whose associated scalar field is also the real number system.

Another familiar system that has the structure of a vector space is the system of all polynomials whose terms are powers of x multiplied by some real number. Typical elements in this system look like this:

$$5x^4 + 3x^2 - 2x + 7$$

$$\sqrt{2}\, x^2 - 3x + 9$$

The two polynomials shown here are of the fourth and second-degree respectively. There are also polynomials that do not contain x at all. They are just ordinary real numbers like 5, 7, -4, $\sqrt{3}$, and so on. We call them polynomials of zero degree. In this system, each polynomial is a vector. The real number system is the associated field of scalars. Vector addition is carried out by adding polynomials according to the rules taught in elementary algebra. The zero element for this addition is 0, a polynomial of no degree. Scalar multiplication is carried out by multiplying a polynomial by a real number according to the rules taught in elementary algebra. With these two operations, the system of polynomials meets all the requirements for being a vector space. In this vector space, each distinct power of x is a separate component. Any one polynomial contains only a finite number of components. But, since any positive power of x, like x^9, or x^{187}, may be used in a polynomial, there is an infinite number of components to choose from in constructing a polynomial. Therefore this system is an example of a vector space of *infinite dimension.*

The *i*, *j*, *k* Notation

In the study of physics it is found that such things as forces, velocities, and rotations can be represented as vectors in a three dimensional vector space. To carry out computations in this vector space the physicist uses a special notation in which all vectors are expressed in terms of three unit vectors called i, j, and k. They are defined as follows: $i = (1, 0, 0)$; $j = (0, 1, 0)$; $k = (0, 0, 1)$. The vector $(2, 3, 5)$ can be expressed in terms of these units in this way: $(2, 3, 5) = (2, 0, 0) + (0, 3, 0) + (0, 0, 5) = 2 \cdot (1, 0, 0) + 3 \cdot (0, 1, 0) + 5 \cdot (0, 0, 1) = 2i + 3j + 5k$.

In general, the vector $(a, b, c) = ai + bj + ck$. When this notation is used, all steps in vector addition and scalar multiplication become simple exercises in elementary high school algebra. For example:

Vector addition:

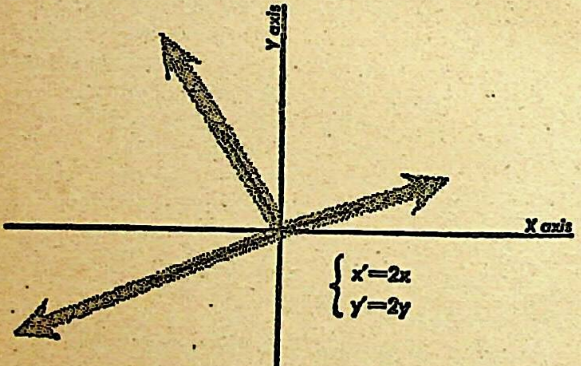
$$(2i + 3j + 5k) + (3i - 4j + 2k) = 5i - j + 7k$$

Scalar multiplication:

$$2(2i + 3j + 5k) = 4i + 6j + 10k$$

Mapping the Plane Into Itself

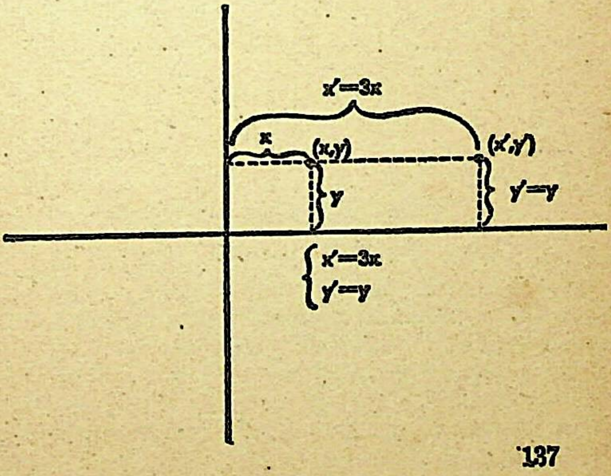
On page 131, we saw that multiplying the vector (2, 1) by the scalar 2 has the effect of doubling the length of the associated arrow without changing its direction. If we multiply each of the vectors in the plane by 2, every one of the associated arrows has its length doubled. The effect of the multiplication is to stretch the entire plane uniformly in all directions, so that each point of the plane is pulled to a new position, twice as far from the origin as it was before. We can represent a typical vector by the symbol (x, y), where x is its x component, and y is its y component. As a

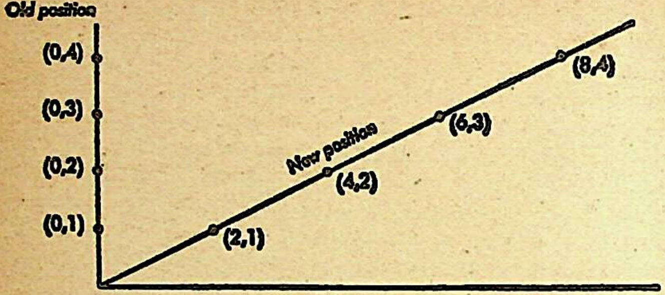


Solid arrow represents (x,y)
Shaded arrow represents (x',y')

result of the stretching, the vector (x, y) is changed into the vector $(2x, 2y)$. If we call the components of the new vector x' and y', the relationship between the old components and the new components is expressed in the equations: $x' = 2x$, and $y' = 2y$.

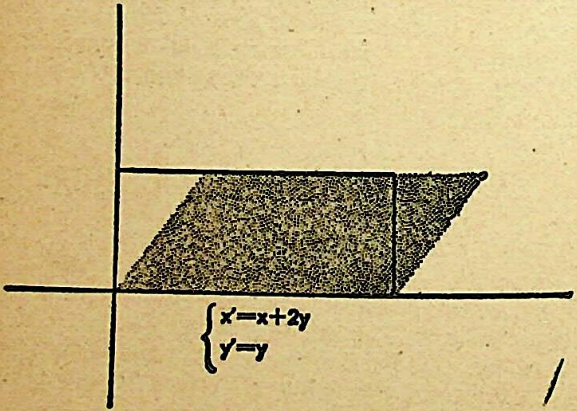
The stretching of the plane is an example of a mapping of the vector space into itself, whereby each element in the vector space is mapped into a particular image. Such a mapping is called a *transformation*. The equations above serve to define the transformation precisely, by giving directions for calculating, from the components of any vector, the components of the image into which it is mapped. If the stretching of the plane is one in which each arrow has its length tripled, the equations are $x' = 3x$ and $y' = 3y$. If we contract the plane, to shrink each arrow to half its size, the equations are $x' = \frac{1}{2}x$, and $y' = \frac{1}{2}y$. If we reverse the direction of each arrow in the vector space, the equations are $x' = -x$, and $y' = -y$. Such a reversal of direction is called a reflection. In general, equations of the form $x' = kx$ and $y' = ky$, where k is some fixed real number, define a stretching or contraction of all vectors, with or without a reversal of direction.





$$\begin{cases} x' = x + 2y \\ y' = y \end{cases}$$

There are other transformations, too, that can be described pictorially in terms of motions or deformations of the plane. Each is defined precisely by an associated pair of equations that shows how the new components (of the image vectors under the mapping) are computed from the old components. The equations $x' = 3x$, $y' = y$ define a transformation which stretches the plane horizontally, while leaving each point as close to the x axis as it was before. The equations $x' = x + 2y$, $y' = y$ define a transformation which also leaves points at the same height above



or below the x axis, while moving them horizontally through varying distances that depend on their height. As a result this mapping moves the points of the y axis into new positions on a sloping line, as shown in the first diagram. The effect is a shearing of the plane, which deforms rectangles into parallelograms, as shown in the second diagram. The equations $x' = x \cos \theta - y \sin \theta$, $y' = x \sin \theta + y \cos \theta$ define a rotation of the plane around the origin as center. All these mappings are special examples of a particular family of transformations whose equations have this simple form:

$$x' = a_1x + a_2y$$
$$y' = b_1x + b_2y$$

Mappings that can be written in this form, where a_1, a_2, b_1, b_2 may be any fixed real numbers, are called *linear mappings.*

Another kind of mapping is obtained when the entire plane is moved in the direction of some particular vector through a distance equal to the length of that vector. Such a mapping is called a *translation.* For example, the equations $x' = x + 2$, $y' = y + 3$ define a translation that moves every point in the plane in the direction and through the distance specified by the vector (2, 3). In general, equations of the form $x' = x + h$, $y' = y + k$, where h and k are fixed real numbers, represent translations of the plane. A more complicated type of transformation is obtained when a linear mapping is followed by a translation. Then the equations look like this:

$$x' = a_1x + a_2y + h$$
$$y' = b_1x + b_2y + k$$

Reversible Mappings

The mapping defined by the equations $x' = x + 2$, $y' = y + 3$ moves every point in the plane two units to the right and three units up. This mapping is easily reversed. If, after the mapping, each point is moved two units to the left and three units down, the points all return to their original positions. The reversibility of the mapping can be expressed in terms of the equations, too. The equations show how to calculate the new components x' and y' from the old components x and y. If we solve the equations for x and y, we get equations in this form: $x = x' - 2$, $y = y' - 3$. These equations give directions for calculating the old components from the new, so they are the equations of the reverse mapping. In this reverse mapping, considered as a mapping in its own right, x' and y' are the components of a vector in its initial position, and x and y are the components in the final position. To conform to the original notation, in which the primed symbols are used to represent the new position, we represent the reverse mapping in this form: $x' = x - 2$, $y' = y - 3$. In general, for every translation, represented by the equations $x' = x + h$, $y' = y + k$ there is a reverse translation whose equations are $x' = x - h$, $y' = y - k$.

The Translations Form a Group

In the equations $x' = x + h$, $y' = y + k$, every time we pick definite values for h and k, we have equations representing some definite translation. For example, the pair of equations $x' = x + 2$, $y' = y + 3$ represents one translation. Let us call this translation P. The pair of equations $x' = x + 5$, $y' = y + 7$ represents another translation. Let us call it Q. If we take all possible values for the numbers h and k, we get all possible translations of the plane into itself. These translations form a system of elements in which an operation of multiplication can be defined.

Just as we did in the case of the symmetries of the triangle, back in Chapter III, we define the product of two translations as the result of performing one translation after another, with the second one taking over where the first one leaves off. The combined effect of the two translations is equivalent to a single transformation, which, in fact, is also a translation.

For example, suppose we designate by $Q * P$ the transformation we get when P is performed first, and Q is performed next on the results of P. P transforms components x and y by adding 2 and 3 to them respectively, so the new components are $x + 2$, and $y + 3$. Q transforms these components by adding 5 and 7 to them, respectively, so the final components are $x + 7$ and $y + 10$. So the equations for the product $Q * P$ are: $x' = x + 7$, $y' = y + 10$. These are of the form $x' = x + h$, $y' = y + k$, with the special values 7 and 10 for h and k. So $Q * P$ also belongs to the system of translations. Because the product of two translations is itself a translation, the operation $*$ is a binary operation in the system of translations. The operation is associative, because, if P, Q, and R are three translations, $(R * Q) * P$ and $R * (Q * P)$ both mean P followed by Q followed by R. The translation defined by the equations $x' = x + 0$, $y' = y + 0$ doesn't move the plane at all, and is the identity element for the operation $*$. If we designate it by I, then $P * I = I * P = P$. We have already observed that every translation is reversible. Let us designate by P^{-1} the translation that is the reverse of P. If it is applied after P, it brings every point in the plane back to its original position. So the product $P^{-1} * P$ is equal to the identity element I. Similarly, $P * P^{-1} = I$. So P^{-1} is the inverse of P with respect to the operation $*$. Because of these properties, the system of translations constitutes a group with respect to the operation $*$. In fact, it is an abelian group, because the operation turns out to be commutative.

The Linear Group

In the equations

$$x' = a_1x + a_2y$$
$$y' = b_1x + b_2y,$$

every time we pick definite values for a_1, a_2, b_1 and b_2, we get equations representing a particular linear mapping of the plane into itself. If we take all possible real values of a_1, a_2, b_1, and b_2, we get all possible linear mappings. They form a system of elements in which the operation $*$, defined in the same way as for translations, is a binary operation, because the product of two linear mappings is itself a linear mapping. For example, if P stands for the mapping

$$x' = 2x + 3y$$
$$y' = 5x - y$$

and Q stands for the mapping

$$x'' = 3x' + 4y',$$
$$y'' = 2x' + 7y'$$

we get the mapping $Q * P$ by applying the mapping Q to the results of P as follows:

$$x'' = 3(2x + 3y) + 4(5x - y) = 6x + 9y + 20x - 4y$$
$$y'' = 2(2x + 3y) + 7(5x - y) = 4x + 6y + 35x - 7y.$$

So the transformation $Q * P$ has the equations

$$x'' = 26x + 5y$$
$$y'' = 39x - y$$

These are the equations of a linear mapping in which a_1, a_2, b_1, and b_2 have the special values 26, 5, 39 and -1 respectively.

However, the system of all linear mappings does not form a group. It fails to qualify as a group because not all the linear mappings are reversible. The reason for this difficulty can be seen by comparing one of the troublesome linear mappings with a translation. A translation maps each point into an image in such a way that no two points have the same image. We can reverse the translation by carrying each point back to the single point of which it is the image. However, the linear mapping defined by the equations

$$x' = 0x + 0y = 0$$
$$y' = 0x + 0y = 0,$$

behaves differently. It carries all points into the origin. As a result, the origin is the image of not one point, but many points. A many-to-one mapping, as we saw on page 15, is not reversible.

However, there are some linear mappings that are reversible. If we put aside those linear mappings that are not reversible, and keep only those that are reversible, then we get a subset of the system of linear mappings that does constitute a group. In this subset, every linear mapping has an inverse, and all the other requirements for qualifying as a group are satisfied. We can find out which linear mappings are reversible by actually trying to reverse one, and seeing what conditions must be fulfilled to attain success. Suppose we take the equations of a linear mapping in general form:

$$x' = a_1x + a_2y$$
$$y' = b_1x + b_2y.$$

To reverse the mapping means to solve for x and y in terms of x' and y'. Let us first write the equations with the x' and y' on the right hand side, and then solve for x and y by the usual method for solving simultaneous equations by eliminating one of the unknowns. To eliminate y, we multiply the first equation by b_2, multiply the second equation by $-a_2$, and then add the resulting equations:

$$\begin{array}{r} a_1b_2x + a_2b_2y = b_2x' \\ -a_2b_1x - a_2b_2y = -a_2y' \\ \hline a_1b_2x - a_2b_1x = b_2x' - a_2y' \end{array}$$

The distributive law permits us to rewrite the left hand side of the equation in factored form: $(a_1b_2 - a_2b_1)x = b_2x' - a_2y'$. The next step in solving for x would be to divide both sides of the equation by $(a_1b_2 - a_2b_1)$ to get as a result

$$x = \frac{b_2x' - a_2y'}{a_1b_2 - a_2b_1}$$

However, this step is not always possible. We know from our discussion on page 74 that division makes sense only when the divisor is not zero. So we can succeed in reversing the linear mapping if and only if $a_1b_2 - a_2b_1$ is not equal to zero. The number $a_1b_2 - a_2b_1$ is called the determinant of the linear mapping. So we can say that a linear mapping is reversible if and only if its determinant is different from zero. For example, suppose linear mappings A and B are defined as follows:

$$A: \begin{cases} x' = 2x + 3y \\ y' = 4x + 6y \end{cases} \qquad B: \begin{cases} x' = 2x - 3y \\ y' = 3x + y \end{cases}$$

The determinant of mapping A is $2 \cdot 6 - 3 \cdot 4 = 0$, so A is not a reversible mapping. The determinant of mapping B is $2 \cdot 1 - (-3) \cdot 3 = 11$, which is different from zero, so B is a reversible mapping. In fact, if we solve for x and y in terms of x' and y', we get

$$x = \frac{1}{11}x' + \frac{3}{11}y'$$

$$y = \frac{-3}{11}x' + \frac{2}{11}y'.$$

These equations have the right form to qualify as a linear mapping, which we may designate as B^{-1}, or the inverse of B. Moreover, B^{-1} is reversible, because solving for x' and y' leads back again to the original equations of B.

The system of reversible linear mappings is known as the *linear group*. It includes all rotations, stretches, and shears of the plane.

If a reversible linear mapping is followed by a translation, the product is called an *affine* transformation. The system of all such products also turns out to be a group, and is known as the affine group. If a uniform stretching of the plane is followed by a translation, the product is called a *similitude*. The system of all such products constitutes another group, known as the group of similitudes. If we put together in one set of transformations all rotations, translations, and reflections, this set, too, is a group, and is known as the *Euclidean* group.

The fact that transformations of a plane into itself can be associated with each other in families of transformations, some of which have a group structure, has led to a new insight into the meaning of geometry. In the geometry we study in high school, considerable time is devoted to the study of congruent figures. We try to find out what figures are congruent to each other. We also investigate properties of a figure that it has in common with any other figure to which it is congruent. Such properties include lengths of corresponding lines, sizes of corresponding angles, area, etc. Two figures were defined as being congruent if they could be made to coincide. This definition implied the use of a motion to carry one figure onto the other. To assure ourselves that the figure would not be deformed while it was being moved, we calmed our fears with the "axiom" that a geometric figure can be moved from place to place without changing its form or size. The effect of this axiom was to banish from the realm of legitimate motions all stretches and shears. At the same time, it singled out as the only legitimate motions those that we have called rotations, translations and reflections. But these are the motions that make up the Euclidean group of transformations. This fact makes it possible to define more precisely what is meant by congruence. Two figures are congruent if one can be mapped onto the other by a transformation that belongs to the Euclidean group. This definition also gives a new meaning to such concepts as length, area, and so on. They turn out to be among the characteristics of a figure that remain unchanged when it is transformed by a mapping that belongs to the Euclidean group.

Another subject treated in high school geometry is that of similar figures. The concept of similarity can also be defined in terms of a group of transformations. Two figures are similar if one can be mapped onto the other by means of a transformation that belongs to the group of similitudes. A similitude does not leave the length of a line unchanged, but it does leave unchanged such things as angles, and ratios of lengths. The fact that the two traditional concerns of plane geometry can be described best in terms of groups of transformations has led to the modern notion of what a geometry is. A geometry is now defined as the study of figures which can be mapped into each other by a group of transformations, and of the properties of figures that remain unchanged when the transformations in the group are applied. In the sense of this definition, what we studied in high school was not geometry, but *some geometries*. When we studied congruence, we were studying Euclidean geometry, associated with the Euclidean group. When we studied similarity, we were studying a different geometry, associated with the group of similitudes. Moreover, there are other geometries which we did not study in high school at all. For example, there is an *affine geometry* associated with the affine group. Because there are many groups that may operate on the same vector space, there is a multiplicity of geometries belonging to one and the same space.

DO IT YOURSELF

1. Use the definition of vector addition given on page 130 to find the following vector sums:

   $(3, 2) + (-1, 2)$ $\qquad$ $(8, -5) + (-5, 8)$

   $(4, 7) + (-4, -7)$ $\qquad$ $(2, 0) + (0, 3)$

2. Use the definition of vector addition to prove that $(a, b) + (c, d) = (c, d) + (a, b)$. (Commutative Law of Addition)
3. Locate the points $(2, 3)$, $(-5, 4)$, and $(-6, -2)$ on a graph with coordinates measured from an x axis and a y axis that are perpendicular to each other. Carry out the following scalar multiplications, and locate the products on the graph: $3 \cdot (2, 3)$; $2 \cdot (-5, 4)$; $-\frac{1}{2} \cdot (-6, -2)$. Verify from the graph that the first two scalar multiplications change the length of the arrow belonging to the vector, without changing its direction. Verify that the

third one changes the length and reverses the direction.

4. Represent these three-dimensional vectors in the i, j, k notation:

$$(4, 6, 1) \qquad (-2, 1, 0) \qquad (0, 5, 7)$$

5. Two linear mappings P and Q (mapping a plane into itself) are defined as follows. Find the equations that define the product $Q * P$.

$$P: \begin{cases} x' = x + y \\ y' = x - y \end{cases} \qquad Q: \begin{cases} x' = 2x + y \\ y' = x - 3y \end{cases}$$

Find the equations that define the product $P * Q$. Is $Q * P = P * Q$? Is the operation $*$ for such linear mappings commutative?

6. Find the inverse of the mapping P defined in question 5.

7. Express as an ordered pair:

a) $(2, 4) + (x, 6)$. b) $(3, -1) + (7, 2)$.
c) $(6, -4) + (-2, 3) + (1, 1)$. d) $4(2, -1) + 3(4, -2)$.
e) $2(-3, 1) + 4(5, -2)$. f) $t(2, 1) + (5, 7)$.

8. Solve the following equations for the vector (x, y):

a) $(x, y) + (2, -3) = (0, 0)$.
b) $(x, y) + (4, 1) = (7, 2)$.
c) $2(x, y) + (3, -2) = (4, 1)$.

9. Solve for the scalars x and y:

a) $x(2, 3) + y(4, -1) = (16, 3)$.
b) $x(2, 4) + y(9, 6) = (4, 4)$.
c) $x(1, 0) + y(0, 1) = (0, 0)$.

10. Find the determinant of each of these linear transformations:

$$A: \begin{cases} x' = 2x + 3y \\ y' = 4x + 6y \end{cases} \qquad B: \begin{cases} x' = 3x - 2y \\ y' = 4x - 3y \end{cases}$$

$$C: \begin{cases} x' = 2x + 3y \\ y' = 3x + 5y \end{cases} \qquad D: \begin{cases} x' = x + y \\ y' = 2x - y \end{cases}$$

11. a) Which of the transformations in Ex. 10 have an inverse?

b) Find these inverses. (See p. 143.)

CHAPTER VII

The Rank and File: Matrices

IN THIS chapter we re-examine from another point of view the linear mappings of a plane into itself. In the course of this re-examination we shall get acquainted with another type of mathematical structure that is important. At the same time we shall acquire some equipment that will be useful to us in the next chapter where we finish the job of constructing a number system that supplies a number for every point in the plane.

Some typical linear mappings of the plane are listed below, with a capital letter assigned to each one as its name:

$$P:\begin{cases} x' = 1x + 1y \\ y' = 1x - 1y \end{cases} \qquad Q:\begin{cases} x' = 2x + 1y \\ y' = 1x - 3y \end{cases}$$

$$O:\begin{cases} x' = 0x + 0y \\ y' = 0x + 0y \end{cases} \qquad I:\begin{cases} x' = 1x + 0y \\ y' = 0x + 1y \end{cases}$$

The mapping called O could have been written more briefly in the form $x' = 0$, $y' = 0$. Similarly, the mapping called I could have been written as $x' = x$, $y' = y$. However, showing the zero coefficients explicitly, as we have in the longer way of writing them, has the advantage of stressing the fact that all linear mappings have the same form. In each mapping, the new x and the new y are obtained by adding some multiple of the old x to some multiple of the old y. Two different linear mappings differ from each other only by virtue of the fact that they use different multiples. What distinguishes one linear mapping from another, then, is the set of four coefficients that appear in the equations of the mapping. This fact suggests that we can represent a mapping in an abbreviated notation in which we delete the letters x, y, x' and y', and merely list these four coefficients, arranging them in a square array, just as they are arranged in the equations written above. Such a square array of numbers is called a *matrix*. In this case, because the matrix has two rows and two columns, we call it a 2 by 2 matrix. Every linear mapping of the plane is associated with such a 2 by 2 matrix, and, vice versa, every 2 by 2 matrix of real numbers belongs to some linear mapping of the plane. To show this correspondence, we shall use as the name of a matrix the name of the mapping that it belongs to. Here are the matrices of the mappings P, Q, O, and I:

$$P = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} \quad Q = \begin{vmatrix} 2 & 1 \\ 1 & -3 \end{vmatrix} \quad O = \begin{vmatrix} 0 & 0 \\ 0 & 0 \end{vmatrix} \quad I = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$$

Matrices Form a Vector Space

Now that we have a collection of matrices, we can disregard their origin as sets of coefficients belonging to linear mappings, and may think of them as constituting an independent system of elements. We proceed to give this system a structure in a simple and natural way. We convert it into a vector space by methods like those we used in the last chapter. Just as the vector (2, 3) is made up of two components, each distinguished by the place it occupies in the ordered pair, each matrix is made up of four components, each distinguished by the place it occupies in the square array. So we can define addition of matrices the way we defined addition of ordered pairs. *To add two matrices, add the corresponding components separately.* We can also define a scalar multiplication for matrices, using the field of real numbers as the field of scalars, in the same way that we defined scalar multiplication for ordered pairs. *To multiply a matrix by a scalar, multiply each of its components by* *that scalar*. Examples of addition and scalar multiplication for matrices are shown below:

$$P + Q = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} + \begin{vmatrix} 2 & 1 \\ 1 & -3 \end{vmatrix} = \begin{vmatrix} 1+2 & 1+1 \\ 1+1 & (-1)+(-3) \end{vmatrix} = \begin{vmatrix} 3 & 2 \\ 2 & -4 \end{vmatrix}$$

$$2 \cdot Q = 2 \cdot \begin{vmatrix} 2 & 1 \\ 1 & -3 \end{vmatrix} = \begin{vmatrix} 2(2) & 2(1) \\ 2(1) & 2(-3) \end{vmatrix} = \begin{vmatrix} 4 & 2 \\ 2 & -6 \end{vmatrix}$$

With addition and scalar multiplication defined in this way, the system of 2 by 2 matrices becomes a vector space. To verify this fact, we have to show that it has all the characteristics of a vector space, listed on page 133. First, we observe that it is an abelian group with respect to addition. This is proved below by showing that addition is associative and commutative, that it has a zero element, and that every matrix has a negative:

Addition is associative:

$$\text{Let } P = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \qquad Q = \begin{vmatrix} c_1 & c_2 \\ d_1 & d_2 \end{vmatrix} \qquad R = \begin{vmatrix} e_1 & e_2 \\ f_1 & f_2 \end{vmatrix}$$

$$(P + Q) + R = \begin{vmatrix} a_1 + c_1 & a_2 + c_2 \\ b_1 + d_1 & b_2 + d_2 \end{vmatrix} + \begin{vmatrix} e_1 & e_2 \\ f_1 & f_2 \end{vmatrix}$$

$$\begin{vmatrix} (a_1 + c_1) + e_1 & (a_2 + c_2) + e_2 \\ (b_1 + d_1) + f_1 & (b_2 + d_2) + f_2 \end{vmatrix}$$

$$P + (Q + R) = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} + \begin{vmatrix} c_1 + e_1 & c_2 + e_2 \\ d_1 + f_1 & d_2 + f_2 \end{vmatrix}$$

$$\begin{vmatrix} a_1 + (c_1 + e_1) & a_2 + (c_2 + e_2) \\ b_1 + (d_1 + f_1) & b_2 + (d_2 + f_2) \end{vmatrix}$$

But $(a_1 + c_1) + e_1 = a_1 + (c_1 + e_1)$, etc., because addition of real numbers is associative. Therefore $(P + Q) + R = P + (Q + R)$.

Addition is commutative:

$$P + Q = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} + \begin{vmatrix} c_1 & c_2 \\ d_1 & d_2 \end{vmatrix} = \begin{vmatrix} a_1 + c_1 & a_2 + c_2 \\ b_1 + d_1 & b_2 + d_2 \end{vmatrix}$$

$$Q + P = \begin{vmatrix} c_1 & c_2 \\ d_1 & d_2 \end{vmatrix} + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} c_1 + a_1 & c_2 + a_2 \\ d_1 + b_1 & d_2 + b_2 \end{vmatrix}$$

But $a_1 + c_1 = c_1 + a_1$, etc., because addition of real numbers is commutative. Therefore $P + Q = Q + P$.

There is a zero element:

$$0 + P = \begin{vmatrix} 0 & 0 \\ 0 & 0 \end{vmatrix} + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} 0 + a_1 & 0 + a_2 \\ 0 + b_1 & 0 + b_2 \end{vmatrix}$$

$$= \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = P$$

Therefore the matrix $0 = \begin{vmatrix} 0 & 0 \\ 0 & 0 \end{vmatrix}$ is the zero element for matrix addition, and is called the zero matrix.

Every matrix has a negative:

In fact, the negative of $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$ is $\begin{vmatrix} -a_1 & -a_2 \\ -b_1 & -b_2 \end{vmatrix}$

because $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} + \begin{vmatrix} -a_1 & -a_2 \\ -b_1 & -b_2 \end{vmatrix} = \begin{vmatrix} a_1 + (-a_1) & a_2 + (-a_2) \\ b_1 + (-b_1) & b_2 + (-b_2) \end{vmatrix}$

$$= \begin{vmatrix} 0 & 0 \\ 0 & 0 \end{vmatrix} = \text{the zero matrix}$$

We observe next that scalar multiplication for matrices obeys the two distributive laws, the mixed associative law, and the law that the number 1 serves as a unity element for scalar multiplication:

Distributive laws:

$$r(P + Q) = r\left(\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} + \begin{vmatrix} c_1 & c_2 \\ d_1 & d_2 \end{vmatrix}\right) = r\begin{vmatrix} a_1 + c_1 & a_2 + c_2 \\ b_1 + d_1 & b_2 + d_2 \end{vmatrix}$$

$$= \begin{vmatrix} r(a_1 + c_1) & r(a_2 + c_2) \\ r(b_1 + d_1) & r(b_2 + d_2) \end{vmatrix} = \begin{vmatrix} ra_1 + rc_1 & ra_2 + rc_2 \\ rb_1 + rd_1 & rb_2 + rd_2 \end{vmatrix}$$

$$rP + rQ = r \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} + r \begin{vmatrix} c_1 & c_2 \\ d_1 & d_2 \end{vmatrix}$$

$$= \begin{vmatrix} ra_1 & ra_2 \\ rb_1 & rb_2 \end{vmatrix} + \begin{vmatrix} rc_1 & rc_2 \\ rd_1 & rd_2 \end{vmatrix} = \begin{vmatrix} ra_1 + rc_1 & ra_2 + rc_2 \\ rb_1 + rd_1 & rb_2 + rd_2 \end{vmatrix}$$

Therefore $r(P + Q) = rP + rQ$.

$$(r + s)P = (r + s) \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} (r + s)a_1 & (r + s)a_2 \\ (r + s)b_1 & (r + s)b_2 \end{vmatrix}$$

$$= \begin{vmatrix} ra_1 + sa_1 & ra_2 + sa_2 \\ rb_1 + sb_1 & rb_2 + sb_2 \end{vmatrix}.$$

$$rP + sP = r \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} + s \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$$

$$= \begin{vmatrix} ra_1 & ra_2 \\ rb_1 & rb_2 \end{vmatrix} + \begin{vmatrix} sa_1 & sa_2 \\ sb_1 & sb_2 \end{vmatrix}$$

$$= \begin{vmatrix} ra_1 + sa_1 & ra_2 + sa_2 \\ rb_1 + sb_1 & rb_2 + sb_2 \end{vmatrix}$$

Therefore $(r + s)P = rP + sP$.

Mixed associative law:

$$r(sP) = r \left(s \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \right) = r \begin{vmatrix} sa_1 & sa_2 \\ sb_1 & sb_2 \end{vmatrix}$$

$$\begin{vmatrix} r(sa_1) & r(sa_2) \\ r(sb_1) & r(sb_2) \end{vmatrix} = \begin{vmatrix} rsa_1 & rsa_2 \\ rsb_1 & rsb_2 \end{vmatrix}$$

$$(rs)P = (rs) \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} (rs)a_1 & (rs)a_2 \\ (rs)b_1 & (rs)b_2 \end{vmatrix}$$

$$= \begin{vmatrix} rsa_1 & rsa_2 \\ rsb_1 & rsb_2 \end{vmatrix}.$$

Therefore $r(sP) = (rs)P$.

Unity element for scalar multiplication:

$$1 \cdot P = 1 \cdot \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} 1a_1 & 1a_2 \\ 1b_1 & 1b_2 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = P$$

Multiplication of Matrices

In the last chapter we defined an operation of multiplication for linear mappings. Since each mapping has an associated matrix, we can transfer this operation to the matrices, and in this way define a multiplication operation for matrices. First, let us write down two matrices P and Q, and the mappings that are associated with them:

$$P = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \quad \text{associated mapping} \quad \begin{cases} x'' = a_1x' + a_2y' \\ y'' = b_1x' + b_2y' \end{cases}$$

$$Q = \begin{vmatrix} c_1 & c_2 \\ d_1 & d_2 \end{vmatrix} \quad \begin{cases} x' = c_1x + c_2y \\ y' = d_1x + d_2y \end{cases}$$

To get the product mapping $P * Q$, we first perform the mapping Q, and then perform the mapping P on the new components we get as a result of having used Q.

$$x'' = a_1(c_1x+c_2y)+a_2(d_1x+d_2y) = a_1c_1x+a_1c_2y+a_2d_1x+a_2d_2y$$
$$y'' = b_1(c_1x+c_2y)+b_2(d_1x+d_2y) = b_1c_1x+b_1c_2y+b_2d_1x+b_2d_2y$$

By rearranging the terms to bring the x terms together, and the y terms together, and then factoring out the x and y, we find that the product mapping $P * Q$ takes this form:

$$x'' = (a_1c_1 + a_2d_1)x + (a_1c_2 + a_2d_2)y$$
$$y'' = (b_1c_1 + b_2d_1)x + (b_1c_2 + b_2d_2)y$$

The matrix that belongs to this mapping is

$$\begin{vmatrix} a_1c_1 + a_2d_1 & a_1c_2 + a_2d_2 \\ b_1c_1 + b_2d_1 & b_1c_2 + b_2d_2 \end{vmatrix}.$$

We call it the product matrix, and designate it by PQ. (We omit the * just as in elementary algebra it is customary to omit the multiplication sign.)

Now, by comparing the matrix PQ with the matrices P and Q, we can observe a simple rule by which the multiplication can be carried out without reference to the equations of the mappings. Notice that in the product matrix PQ, the component in the *first row* and *first* column is $a_1c_1 + a_2d_1$. This expression can be obtained from the pairs (a_1, a_2) and (c_1, d_1) by first multiplying corresponding terms in the pairs, and then adding the products. The pair (a_1, a_2) is the *first row* of matrix P. The pair (c_1, d_1) is the *first column* of the matrix Q. So we get the component in the first row and first column of the product PQ by multiplying corresponding components of the first row of P and the first column of Q, and adding the results. In order to be able to refer to it in fewer words, let us designate this kind of operation as multiplication of the first row of P by the first column of Q. We can now use this language to describe the procedure for getting all the components of the product matrix. To get the component in the first row and second column, multiply the first row of P by the second column of Q. To get the component in the second row and first column, multiply the second row of P by the first column of Q. To get the component in the second row and second column, multiply the second row of P by the second column of Q. In general, to get the component of the product matrix PQ that is in any particular row and column, multiply the corresponding row of P by the corresponding column of Q.

If we apply this rule for multiplying matrices to get the product QP, we find that

$$QP = \begin{vmatrix} c_1 & c_2 \\ d_1 & d_2 \end{vmatrix} \cdot \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} c_1a_1 + c_2b_1 & c_1a_2 + c_2b_2 \\ d_1a_1 + d_2b_1 & d_1a_2 + d_2b_2 \end{vmatrix}.$$

Notice that this product is not the same as PQ, so matrix multiplication is *not commutative.*

The Matrices Form a Ring

Although matrix multiplication is not commutative, it is associative. We do not have to give a special proof of this fact, because we have already observed that the operation *, performed on linear mappings, is associative, and matrix multiplication is merely another way of expressing the same operation. It is also possible to show that matrix multiplication is distributive with respect to matrix addition. Consequently, the system of 2 by 2 matrices meets all the requirements for being a ring, just as the natural numbers, the integers, the rational numbers, and the real numbers did. However, unlike these other systems, it is not a commutative ring, since matrix multiplication is not commutative.

There is a unity element for matrix multiplication. In fact, the unity element is the matrix which we have called I. Since multiplication is not, in general, commutative, we have to verify separately that I behaves as a unity element when it is used as a multiplier either from the left or from the right.

$$IP = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \cdot \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} 1a_1 + 0b_1 & 1a_2 + 0b_2 \\ 0a_1 + 1b_1 & 0a_2 + 1b_2 \end{vmatrix}$$

$$= \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = P$$

$$PI = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \cdot \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = \begin{vmatrix} a_1 \cdot 1 + a_2 \cdot 0 & a_1 \cdot 0 + a_2 \cdot 1 \\ b_1 \cdot 1 + b_2 \cdot 0 & b_1 \cdot 0 + b_2 \cdot 1 \end{vmatrix}$$

$$= \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = P$$

Therefore $IP = PI = P$.

It is natural to ask whether the ring of 2 by 2 matrices is also a field. We find that it is not a field, because it contains zero divisors, and a field may not have zero divisors, as we saw on page 87. Recall that zero divisors are non-zero elements that have a zero product. The matrices $\begin{vmatrix} 0 & 2 \\ 0 & 0 \end{vmatrix}$ and $\begin{vmatrix} 0 & 3 \\ 0 & 0 \end{vmatrix}$ are different from the zero matrix, but their product is equal to the zero matrix:

$$\begin{vmatrix} 0 & 2 \\ 0 & 0 \end{vmatrix} \cdot \begin{vmatrix} 0 & 3 \\ 0 & 0 \end{vmatrix} = \begin{vmatrix} 0 \cdot 0 + 2 \cdot 0 & 0 \cdot 3 + 2 \cdot 0 \\ 0 \cdot 0 + 0 \cdot 0 & 0 \cdot 3 + 0 \cdot 0 \end{vmatrix} = \begin{vmatrix} 0 & 0 \\ 0 & 0 \end{vmatrix} = 0$$

In fact, each of these matrices has the peculiar property that when it is multiplied by itself, the product is zero. If we use the letter T to represent the matrix $\begin{vmatrix} 0 & 2 \\ 0 & 0 \end{vmatrix}$, then

$$T^2 = T \cdot T = \begin{vmatrix} 0 & 2 \\ 0 & 0 \end{vmatrix} \cdot \begin{vmatrix} 0 & 2 \\ 0 & 0 \end{vmatrix} = \begin{vmatrix} 0 \cdot 0 + 2 \cdot 0 & 0 \cdot 2 + 2 \cdot 0 \\ 0 \cdot 0 + 0 \cdot 0 & 0 \cdot 2 + 0 \cdot 0 \end{vmatrix}$$

$$= \begin{vmatrix} 0 & 0 \\ 0 & 0 \end{vmatrix} = 0$$

An element like T that has the property that there is a power of it that is equal to zero is called a *nilpotent* element. In the rings we met before, the zero element itself was the only nilpotent element. However, there are many rings which, like the ring of 2 by 2 matrices, have nilpotent elements that are different from zero.

The Matrices Form an Algebra

The ring of 2 by 2 matrices has a double structure. With matrix addition and scalar multiplication, it has the structure of a vector space. With matrix addition and matrix multiplication, it has the structure of a ring. A system that has such a double structure is known as an *algebra*. Roughly, an algebra may be described as a ring that also has a scalar multiplication defined for it, or a vector space that also has a binary operation of multiplication defined for it. The algebra we have just examined is only one of many possible algebras. In fact, by using definitions of addition, scalar multiplication, and matrix multiplication analogous to those given for 2 by 2 matrices, we can construct an algebra of 3 by 3 matrices, an algebra of 4 by 4 matrices, etc.

Rectangular Matrices

So far we have considered as matrices only square arrays of numbers. However, it is possible to extend the concept of a matrix to include rectangular arrays as well. If a rectangular array of numbers has two rows and five columns, for example, it is known as a 2 by 5 matrix. A vector space can be constructed out of all 2 by 5 matrices by defining matrix addition and scalar multiplication in the same way that we did for square matrices. In general, we can give a vector space structure in this way to the system of all rectangular matrices that have any fixed number of rows and any fixed number of columns. From this point of view, the vectors we first described in the last chapter turn out to be special cases of matrices. An ordered pair like (2, 3) is nothing but a 1 by 2 matrix. An ordered triple is a 1 by 3 matrix, and so on. It is also possible to have a 2 by 1 matrix that has two rows and one column, or a 3 by 1 matrix that has three rows and one column.

The definition of matrix multiplication can also be extended to rectangular matrices, with a special restriction that arises from the way in which matrix multiplication is carried out. To multiply two matrices, we have to multiply each of the rows of the first matrix by each of the columns in the second matrix. This is possible only if a row of the first matrix has as many components as a column does in the second matrix. This means that multiplication of matrices is possible only when the number of columns in the first matrix is equal to the number of rows in the second matrix. For example, we can multiply a 2 by 3 matrix times a 3 by 4 matrix, but we cannot multiply a 3 by 4 matrix times a 2 by 3 matrix.

Rectangular matrices give us a very powerful condensed language in which systems of several equations may sometimes be expressed as a single simple matrix equation. For example, consider these equations for a linear transformation that maps the ordered triples (x_1, x_2, x_3) into the ordered triples (y_1, y_2, y_3):

$$y_1 = a_1x_1 + a_2x_2 + a_3x_3$$

$$y_2 = b_1x_1 + b_2x_2 + b_3x_3$$

$$y_3 = c_1x_1 + c_2x_2 + c_3x_3$$

These three equations are equivalent to the single matrix equation:

$$\begin{vmatrix} y_1 \\ y_2 \\ y_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \cdot \begin{vmatrix} x_1 \\ x_2 \\ x_3 \end{vmatrix}$$

as you can verify by carrying out the matrix multiplication. If we introduce names for the matrices, as follows,

$$Y = \begin{vmatrix} y_1 \\ y_2 \\ y_3 \end{vmatrix} \qquad P = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \qquad X = \begin{vmatrix} x_1 \\ x_2 \\ x_3 \end{vmatrix},$$

then the equation takes the particularly simple form, $Y = PX$. Working with this single equation according to the rules of matrix algebra then takes the place of working with the three original equations.

Matrix algebra, one of the youngest branches of mathematics, is now one of the most widely used. Besides being an indispensable tool in higher mathematics, it is also employed in such diverse fields as psychology, chemistry, physics, economics, and electrical engineering.

DO IT YOURSELF

1. Find the sums $P + Q$ and $Q + P$, and the products PQ and QP of the following 2 by 2 matrices:

$$P = \begin{vmatrix} 2 & 1 \\ 4 & 3 \end{vmatrix} \qquad Q = \begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix}$$

Compare your answers to verify that $P + Q = Q + P$, but PQ is not equal to QP.

2. a) Find the scalar product $2\begin{vmatrix} a & b \\ c & d \end{vmatrix}$.

b) Find the matrix product $\begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} \cdot \begin{vmatrix} a & b \\ c & d \end{vmatrix}$.

Compare the results of both multiplications. Notice that the matrix $\begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix}$, used with matrix multiplication, behaves like the scalar 2, used with scalar multiplication.

3. Let $P = \begin{vmatrix} 0 & 1 \\ 2 & 3 \end{vmatrix} \qquad Q = \begin{vmatrix} -1 & 2 \\ 4 & 3 \end{vmatrix} \qquad R = \begin{vmatrix} 2 & -1 \\ 6 & 5 \end{vmatrix}$

a) Find $P(Q + R)$. That is, add Q and R, and then multiply by P from the left.

b) Find $PQ + PR$. That is, find the products PQ and PR, and then add them.

c) Compare the results of a) and b) to verify that the distributive law is obeyed.

4. Let $T = \begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{vmatrix}$

a) Find $T^2 = TT$.

b) Find $T^3 = T^2T$. What kind of element is T? (See page 156)

CHAPTER VIII

Arrows That Are Numbers

NOW WE return to unfinished business. In Chapter VI we had set out to accomplish a double purpose. We wanted to construct an extension of the real number system that would supply a number for every point in a plane, and would at the same time contain a number that satisfies the equation $x^2 = -1$. As the first step toward accomplishing this purpose we constructed a system in which each element is an ordered pair of real numbers, like (1, 4), or $(-\sqrt{2}, \frac{1}{2}\pi)$. We defined an operation of addition for these elements, and an operation we called scalar multiplication. With these two operations the system became an example of the type of structure we called a vector space.

In this vector space, we have an element for every point in the plane. However, this does not mean that we have already accomplished our purpose. Our goal was to find a *number system* that supplies an element for every point in the plane. The vector space we have constructed does not yet qualify as a number system. A structure is entitled to be called a number system only if it has two binary operations called addition and multiplication, such that each of these operations is associative and commutative, and multiplication is distributive with respect to addition. In the vector space whose elements are ordered pairs, we have an addition operation that is associative and commutative. We also have the operation called scalar multiplication, but it does not help the system qualify as a number system, because scalar multiplication is not a binary operation. In scalar multiplication, we multiply a vector by an element that is outside the system of vectors. In a multiplication that is a binary operation, we would have to multiply a vector by a vector, and get a product that is also a vector. So, to complete our construction, we now define a multiplication of this kind.

Ordered Pairs Become Numbers

We define the multiplication of ordered pairs by the following equation: $(a, b) \cdot (c, d) = (ac - bd, ad + bc)$. (In this equation, ac means $a \cdot c$, in accordance with the custom of sometimes omitting the multiplication sign in indicated products.) With this definition, for example, the product $(2, 3) \cdot (4, 1)$ would be equal to $(2 \cdot 4 - 3 \cdot 1, 2 \cdot 1 + 3 \cdot 4) = (5, 14)$. Now we show that this operation is commutative and associative, and is distributive with respect to addition.

To show that multiplication of ordered pairs is commutative, we compare $(a, b) \cdot (c, d)$ with $(c, d) \cdot (a, b)$.

$$(a, b) \cdot (c, d) = (ac - bd, ad + bc)$$

$$(c, d) \cdot (a, b) = (ca - db, cb + da).$$

But, in the real number system, $ac - bd = ca - db$, and $ad + bc = cb + da$. So the two products are the same, and the commutative law is obeyed.

To show that multiplication of ordered pairs is associative, we compare $[(a, b) \cdot (c, d)] \cdot (e, f)$ with $(a, b) \cdot [(c, d) \cdot (e, f)]$.

$$\begin{aligned}[(a, b) \cdot (c, d)] \cdot (e, f) &= (ac - bd, ad + bc) \cdot (e, f)\\ &= ([ac - bd]e - [ad + bc]f, [ac - bd]f + [ad + bc]e)\\ &= (ace - bde - adf - bcf, acf - bdf + ade + bce).\end{aligned}$$

$$\begin{aligned}(a, b) \cdot [(c, d) \cdot (e, f)] &= (a, b) \cdot (ce - df, cf + de)\\ &= (a[ce - df] - b[cf + de], a[cf + de] + b[ce - df])\\ &= (ace - adf - bcf - bde, acf + ade + bce - bdf).\end{aligned}$$

The first component in each of these products is the sum of four terms. Notice that they are the same four terms, merely written in a different order. Therefore these components are equal. Similarly, the second components are equal, and therefore the two products are equal. So multiplication of ordered pairs obeys the associative law.

To show that multiplication of ordered pairs is distributive with respect to addition, we compare $(a, b) \cdot [(c, d) + (e, f)]$ with $(a, b) \cdot (c, d) + (a, b) \cdot (e, f)$.

$$\begin{aligned}(a, b) \cdot [(c, d) + (e, f)] &= (a, b) \cdot (c + e, d + f)\\ &= (a[c + e] - b[d + f], a[d + f] + b[c + e])\\ &= (ac + ae - bd - bf, ad + af + bc + be).\end{aligned}$$

$$\begin{aligned}(a, b) \cdot (c, d) &+ (a, b) \cdot (e, f)\\ &= (ac - bd, ad + bc) + (ae - bf, af + be)\\ &= ([ac - bd] + [ae - bf], [ad + bc] + [af + be])\\ &= (ac - bd + ae - bf, ad + bc + af + be).\end{aligned}$$

The comparison shows that both expressions lead to the same result, and the distributive law is obeyed.

With the addition defined in Chapter VI, and with this new kind of multiplication, the system of ordered pairs of real numbers obeys the five laws that are characteristic of a number system. So now, at last, we have a number system that supplies a number for every point in a plane. We saw in Chapter VI that each ordered pair can be represented pictorially by an arrow, and the operations defined for ordered pairs may be interpreted as operations with the arrows. So we have, in effect, converted the system of arrows into a number system. We call it the system of *complex numbers.*

We have already observed, in Chapter VI, that the system of ordered pairs is an abelian group with respect to addition. Now that we have for this system a binary operation of multiplication that is associative and obeys the distributive law, the system meets the requirements for being a ring. But we already know that it is a vector space, equipped with a scalar multiplication. So the system of complex numbers has a double structure and therefore qualifies as an *algebra* in the sense of the definition given on page 156.

The Complex Number System Is a Field

To qualify as a field as well as a ring, the complex number system must meet the requirements of having a unity element, and having a reciprocal for every element except the zero element. We observe first that the complex number (1, 0) is a unity element for the system, because when it multiplies any number in the system, it leaves that number unchanged:

$$(1, 0) \cdot (a, b) = (1 \cdot a - 0 \cdot b, 1 \cdot b + 0 \cdot a) = (a, b).$$

Now we shall produce a reciprocal for every complex number that is different from the zero element. The zero element in the complex number system is (0, 0). So a complex number (a, b) is the zero element only if $a = 0$ and $b = 0$. If a complex number is different from the zero element, then a and b are not both zero. In that case, $a^2 + b^2$ is not zero, so that division by $a^2 + b^2$ is possible. So, if (a, b) is not the zero element, the symbols $\frac{a}{a^2 + b^2}$ and $-\frac{b}{a^2 + b^2}$ represent actual real numbers, and the ordered pair $\left(\frac{a}{a^2+b^2}, -\frac{b}{a^2+b^2}\right)$ is a complex number. We now show that it is the reciprocal of (a, b) by showing that their product is equal to the unity element:

$$(a, b) \cdot \left(\frac{a}{a^2 + b^2}, -\frac{b}{a^2 + b^2}\right)$$

$$= \left(a\left[\frac{a}{a^2+b^2}\right] - b\left[-\frac{b}{a^2+b^2}\right], a\left[-\frac{b}{a^2+b^2}\right] + b\left[\frac{a}{a^2+b^2}\right]\right)$$

$$= \left(\frac{a^2}{a^2 + b^2} + \frac{b^2}{a^2 + b^2}, -\frac{ab}{a^2 + b^2} + \frac{ab}{a^2 + b^2}\right)$$

$$= \left(\frac{a^2 + b^2}{a^2 + b^2}, 0\right) = (1, 0).$$

For example, the reciprocal of $(3, 4)$ is $(\frac{3}{25}, -\frac{4}{25})$.

This fact can be verified separately by multiplication:

$$(3, 4) \cdot (\tfrac{3}{25}, -\tfrac{4}{25}) = (3 \cdot [\tfrac{3}{25}] - 4[-\tfrac{4}{25}], 3[-\tfrac{4}{25}] + 4[\tfrac{3}{25}])$$

$$= (\tfrac{9}{25} + \tfrac{16}{25}, -\tfrac{12}{25} + \tfrac{12}{25}) = (1, 0)$$

We Still Have the Real Numbers

When we constructed the system of integers, we showed that there is a subset of the integers, the positive integers, that is isomorphic to the natural number system, and therefore can take its place for all practical purposes. In this sense, the system of integers includes the system of natural numbers, and is therefore an extension of the natural number system. In the same way we showed that the rational number system includes a subset isomorphic to the integers, and the real number system includes a subset isomorphic to the rational numbers. Now we show that the complex number system includes a subset isomorphic to the real numbers, and is therefore an extension of the real number system.

Before we display this subset, let us first introduce a new notation analogous to the i, j, k notation used for three dimensional vectors in Chapter VI. The complex number system is a two dimensional vector space, and all of its elements can be expressed in terms of the two unit vectors $(1, 0)$ and $(0, 1)$. We have already seen that $(1, 0)$ is the unity element of the system, so let us call it u, to remind us of that fact. We assign the name i to the vector $(0, 1)$. Then, the complex number $(a, b) = (a, 0) + (0, b) = a \cdot (1, 0) + b \cdot (0, 1) = au + bi$. These equalities follow from the rules for vector addition and scalar multiplication. Now we shall show that the real number system is isomorphic to the subset of the complex number system consisting of those ordered pairs in which the second component is 0. All the numbers in this subset have the special form $(a, 0) = a \cdot (1, 0) = au$. To prove the isomorphism we have to produce a one-to-one correspondence in which the image of a sum is the sum of the images, and the image of a product is the product of the images.

The one-to-one correspondence we use is the mapping $a \leftrightarrow au$. That is, we associate with each real number a the complex number that is equal to a times the unity element u. Under this mapping, the image of 1 is $1u$, the image of 5 is $5u$, the image of $\frac{1}{2}$ is $\frac{1}{2}u$, the image of $\sqrt{3}$ is $\sqrt{3}u$, the image of -2 is $-2u$. The image of 0 is $0u = 0(1, 0) = (0, 0) =$ the zero element of the complex number system. The image of $a + b$ is $(a + b)u$, and the image of ab is $(ab)u$.

Now let us compare the image of a sum with the sum of the images. The image of a is au. The image of b is bu. The sum of the images is $au + bu = a(1, 0) + b(1, 0) = (a, 0) + (b, 0) = (a + b, 0) = (a + b)(1, 0) = (a + b)u$. But this is precisely the image of the sum $a + b$ under the mapping. So the image of a sum is the sum of the images, and the mapping preserves addition.

Now let us compare the image of a product with the product of the images. The product of the images is $(au) \cdot (bu) = [a(1, 0)] \cdot [b(1, 0)] = (a, 0) \cdot (b, 0) = (ab - 0 \cdot 0, a \cdot 0 + b \cdot 0) = (ab, 0) = (ab)(1, 0) = (ab)u$. But this is precisely the image of the product ab. So the image of a product is the product of the images, and the mapping preserves multiplication. Therefore, in so far as the binary operations of addition and multiplication are concerned, the complex number au behaves just like the real number a, and, in particular, the unity element u behaves just like the real number 1. However, we have one more comparison to make, before we can agree that real numbers and complex numbers of the form au are interchangeable. The real number system does a special job in relation to the complex number system when it serves as the field of scalars for scalar multiplication. We have to check whether complex numbers of the form au can serve as scalars, too. To answer this question, we multiply any complex number (c, d) by the scalar a, using scalar multiplication, of course. Then we multiply the same complex number (c, d) by the complex image of a, namely au. Since both (c, d) and au

are complex numbers, this multiplication will be the binary operation defined for the complex number system. Then we compare the results.

Scalar multiplication: $a(c, d) = (ac, ad)$.

Complex number multiplication:

$$(au) \cdot (c, d) = [a(1, 0)] \cdot (c, d) = (a, 0) \cdot (c, d)$$
$$= (ac - 0 \cdot d, ad + 0 \cdot c) = (ac, ad).$$

The products are the same! Multiplication by the complex number au produces the same result as scalar multiplication by the real number a. So the complex numbers of the form au and the real numbers are completely equivalent, and the complex number system is an extension of the real number system. We can now dispense with the special symbol u, and replace it by 1, the symbol for the real number that it is equivalent to. Similarly, we write a for the complex number au. So now, the complex number (a, b), which we have written as $au + bi$, may be written as $a + bi$. The a is referred to as its real part and the bi is referred to as its *imaginary part*, and the complex number i is called the imaginary unit. In this notation, a real number is a complex number $a + bi$ whose imaginary part has $b = 0$.

The fact that scalar multiplication by the real number a is equivalent to multiplication by the complex number au (now also designated by a) makes it unnecessary to retain scalar multiplication as a separate operation for the algebra of complex numbers. Any time we need scalar multiplication, we merely think of it as the special case of the multiplication of complex numbers that arises when one of the multipliers happens to be a real number.

The Question Is Answered

One of our purposes in constructing the system of complex numbers was to have a system in which the equation

$x^2 = -1$ has a solution. We can now show that this purpose has been achieved. First let us be sure that we can recognize -1 in this system when we see it. The number -1, as an element in the complex number system, is equivalent to what we called $-1u$ before, which is $-1(1, 0)$, or $(-1, 0)$. It is in this form that we shall encounter it. The equation that we are trying to solve asks the question, "What number multiplied by itself gives -1 as the product?" We now show that the imaginary unit i provides the answer to the question.

$$\begin{aligned} i \cdot i = (0, 1) \cdot (0, 1) &= (0 \cdot 0 - 1 \cdot 1, 0 \cdot 1 + 1 \cdot 0) \\ &= (-1, 0) = -1. \end{aligned}$$

Now that we have established the fact that $i^2 = i \cdot i = -1$, we can discard the cumbersome apparatus of ordered pairs, and work with complex numbers in a particularly easy way. We write them in the form $a + bi$, as we have already done, and we add them and multiply them according to the rules of elementary high school algebra, making use of the special rule that $i^2 = -1$. How addition and multiplication of complex numbers are carried out in this notation is shown in the following examples:

To add $(2 + 3i) + (-5 + 6i)$:

$$\begin{array}{r} 2 + 3i \\ \underline{-5 + 6i} \\ -3 + 9i \end{array}$$

To multiply $(2 + 3i)(-5 + 6i)$:

$$\begin{array}{rrr} 2 + 3i & & \\ \underline{-5 + 6i} & & \\ -10 - 15i & & \\ \underline{ + 12i} & \underline{+ 18i^2} & \\ -10 - 3i & + 18(-1) & \\ = -28 - 3i & & \end{array}$$

Bad Names for Good Numbers

Complex numbers were first introduced by the Italian mathematicians of the sixteenth century, who found, as we have, that they could not solve certain equations without them. However, mathematicians in those days were accustomed to think in terms of only one number system of positive and negative numbers. They considered the numbers in this system genuine, and therefore called them "real." The complex numbers clearly did not belong to the "real" number system, so, although they used them as a convenience, they considered them to be spurious, or unreal, and called them "imaginary." In fact, the great seventeenth century mathematician, René Descartes, was so doubtful of the reality of these numbers that he rejected them altogether. Today we realize that there is not only one number system. There are many number systems. All are equally genuine, although they differ from each other. Nevertheless we still use the old names, "real" numbers and "imaginary" numbers. So we must be careful not to be influenced by the old prejudices that are expressed in these names. When we use the term "imaginary numbers" now, we must bear in mind that it is a technical term, and should not be interpreted as a derogatory epithet casting doubt on the genuineness of the numbers.

Imaginary numbers are genuine numbers. We have demonstrated their existence by constructing them in the form of ordered pairs of real numbers. However, in order to dispel any lingering doubts that may remain about their genuineness, we now proceed to construct them by two other methods. After constructing complex numbers in three different ways, we ought to feel certain that they are genuine, though not "real."

Complex Numbers as Matrices

The second construction of the complex number system requires no new act of creation on our part. We find that the complex numbers are actually hidden in one of the structures we built up before. They form a subset of the algebra of 2 by 2 matrices defined in Chapter VII. In fact, the complex numbers are nothing else but the 2 by 2 matrices of the form $\begin{vmatrix} a & b \\ -b & a \end{vmatrix}$, in which the components in the upper left hand corner and the lower right hand corner are the same real number, and the other two components have the property that each is the negative of the other. Examples of this type of matrix are:

$$\begin{vmatrix} 2 & 3 \\ -3 & 2 \end{vmatrix} \quad \begin{vmatrix} -6 & -5 \\ 5 & -6 \end{vmatrix} \quad \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \quad \begin{vmatrix} 0 & 0 \\ 0 & 0 \end{vmatrix} \quad \begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix}.$$

It may seem strange for this to be true, in view of the fact that multiplication of matrices is not commutative, while multiplication of complex numbers is commutative. However, while multiplication may not be commutative for a system as a whole, it may be commutative for some special subset in the system. In any case, we can easily prove that multiplication is commutative for the 2 by 2 matrices that have this special form. We compare

$$\begin{vmatrix} a & b \\ -b & a \end{vmatrix} \cdot \begin{vmatrix} c & d \\ -d & c \end{vmatrix} \text{ with } \begin{vmatrix} c & d \\ -d & c \end{vmatrix} \cdot \begin{vmatrix} a & b \\ -b & a \end{vmatrix},$$

and show that the products are the same,

$$\begin{vmatrix} a & b \\ -b & a \end{vmatrix} \cdot \begin{vmatrix} c & d \\ -d & c \end{vmatrix} = \begin{vmatrix} a(c) + b(-d) & a(d) + b(c) \\ -b(c) + a(-d) & -b(d) + a(c) \end{vmatrix}$$

$$= \begin{vmatrix} ac - bd & ad + bc \\ -bc - ad & -bd + ac \end{vmatrix}.$$

$$\begin{vmatrix} c & d \\ -d & c \end{vmatrix} \cdot \begin{vmatrix} a & b \\ -b & a \end{vmatrix} = \begin{vmatrix} c(a) + d(-b) & c(b) + d(a) \\ -d(a) + c(-b) & -d(b) + c(a) \end{vmatrix}$$

$$= \begin{vmatrix} ac - bd & bc + ad \\ -ad - bc & -bd + ac \end{vmatrix}.$$

An examination of the two products, component by component, shows that they are the same.

Actually, it is not necessary to give this special proof that multiplication of these matrices is commutative. All we really have to do is show that the system of these special matrices is isomorphic with the complex number system we already have. Once we have established this fact, then it follows that the system of special matrices has all the properties that the complex number system has, including the commutativity of multiplication.

To prove that the two systems are isomorphic, we make use of the mapping $a + bi = (a, b) \longleftrightarrow \begin{vmatrix} a & b \\ -b & a \end{vmatrix}$. Under this mapping, $(0, 0)$ is associated with $\begin{vmatrix} 0 & 0 \\ 0 & 0 \end{vmatrix}$, $(1, 0)$ is associated with $\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$, and $(0, 1)$ is associated with $\begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix}$. This mapping is clearly a one-to-one correspondence. Now we show that it preserves addition and multiplication. The ordered pairs (a, b) and (c, d) have the sum $(a + c, b + d)$. Their images in the system of matrices are $\begin{vmatrix} a & b \\ -b & a \end{vmatrix}$ and $\begin{vmatrix} c & d \\ -d & c \end{vmatrix}$. These images have the matrix sum $\begin{vmatrix} a+c & b+d \\ -b-d & a+c \end{vmatrix}$, which is the image of the ordered pair $(a + c, b + d)$ under the mapping. So the sum of the images is the image of the sum.

The ordered pairs (a, b) and (c, d) have the product $(ac - bd, ad + bc)$. Their images, $\begin{vmatrix} a & b \\ -b & a \end{vmatrix}$ and $\begin{vmatrix} c & d \\ -d & c \end{vmatrix}$, have the matrix product $\begin{vmatrix} ac - bd & ad + bc \\ -bc - ad & -bd + ac \end{vmatrix}$, as shown on page 169. But this product is the image of the ordered pair $(ac - bd, ad + bc)$ under the mapping. So the product of the images is the image of the product. Therefore this system of special 2 by 2 matrices is isomorphic to the complex number system. It is nothing but the complex number system masquerading in a different style of dress. In this representation, the number 1 appears as $\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$. The real numbers have the form $\begin{vmatrix} a & 0 \\ 0 & a \end{vmatrix}$. The numbers bi, whose real part is zero, have the form $\begin{vmatrix} 0 & b \\ -b & 0 \end{vmatrix}$. Consequently, the number i is $\begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix}$, and the number -1 is $\begin{vmatrix} -1 & 0 \\ 0 & -1 \end{vmatrix}$.

We can verify directly by matrix multiplication that the number i in this form satisfies the equation $x^2 = -1$:

$$i^2 = \begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix} \cdot \begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix} = \begin{vmatrix} 0 \cdot 0 + 1(-1) & 0 \cdot 1 + 1 \cdot 0 \\ -1 \cdot 0 + 0(-1) & -1 \cdot 1 + 0 \cdot 0 \end{vmatrix}$$

$$= \begin{vmatrix} -1 & 0 \\ 0 & -1 \end{vmatrix} = -1.$$

Complex Numbers as Residue Classes

The third method of constructing the complex number system follows a procedure we have already used before. In Chapter III we divided the system of integers into residue classes modulo 3, by putting together into the same class all integers that have the same remainder when we divide them by 3. We defined operations of addition and multiplication for these residue classes. Then we found that, with these operations, the residue classes formed a number system. In Chapter IV we found out that the system of residue classes modulo 3 is also a field. We are now going to go through these steps in the same order. First, we shall construct a ring that will play the same part here that the ring of integers did in Chapter III. Then we shall divide it into residue classes, by putting into the same class all members of the ring that have the same remainder when we divide by a particular member of the ring specially chosen for our purpose. Then, with addition and multiplication of residue classes defined as it was in Chapter III we shall find that the system of residue classes is a number system, and a field. Finally, we shall show that this number system is nothing but the complex number system in disguise.

The elements that we use for the ring that we start with are the polynomials whose terms are powers of x with coefficients that are real numbers. In this system, each separate real number, like ½, or 2, or $\sqrt{3}$, is a polynomial of zero degree. Then there are polynomials of the first degree, like $2x + 5$, and $\sqrt{3}x - 7$. There are polynomials of the second degree, like $x^2 - 1$, $x^2 + 1$, and polynomials of higher degree like $2x^3 - 5x^2 + 6x + 7$, $x^5 - 4x + 1$, etc. On page 134, we saw that this system of polynomials is a vector space. The addition in this vector space is the ordinary addition of polynomials we learned in high school algebra. The polynomials form an abelian group with respect to this operation of addition. The zero element for this operation is the real number 0, considered as a polynomial of no degree. Now we introduce an operation of multiplication of polynomials by using the rules of multiplication taught in elementary high school algebra. For example, we obtain the product $(2x + 1)(3x - 5)$ in this way:

$$\begin{array}{rrr}
3x & -\ 5 & \\
2x & +\ 1 & \\
\hline
6x^2 & -\ 10x & \\
 & +\ 3x & -\ 5 \\
\hline
6x^2 & -\ 7x & -\ 5
\end{array}$$

This operation of multiplication of polynomials is associative and commutative, and it is distributive with respect to addition. So the system of polynomials has the structure of a ring. The unity element for multiplication of polynomials is the real number 1 considered as a polynomial of zero degree.

The next step we take is influenced by the purpose we are trying to accomplish. Our goal, you will recall, is to construct a number system in which the equation $x^2 + 1 = 0$ has a solution. The polynomial that appears in this equation plays a special part in the next step of our construction. We use it as a divisor, and divide it into every other polynomial in the system. Where the division comes out even, we get a quotient, and a remainder equal to zero. Where the division does not come out even, the remainder is different from zero. For example, when we divide $x^4 + 3x^2 + 2$ by $x^2 + 1$, the quotient is $x^2 + 2$, and the remainder is 0. When we divide $x^2 + 1$ by $x^2 + 1$, the quotient is 1, and the remainder is 0. When we divide 0 by $x^2 + 1$, the quotient is 0, and the remainder is 0. When we divide $x^2 + 2$ by $x^2 + 1$, the quotient is 1 and the remainder is 1. When we divide x by $x^2 + 1$, the quotient is 0, and the remainder is x. When we divide $x^2 + 3x + 5$ by $x^2 + 1$, the quotient is 1, and the remainder is $3x + 4$.

In arithmetic, when we divide one integer by another, the remainder is always smaller than the divisor. There is an analogous rule for the division of polynomials. When we divide one polynomial by another, the remainder is always a polynomial of lower degree than the divisor. This is so because, as long as the remainder has the same degree as the divisor, or a higher degree, the division can be carried out for at least one more step. This is seen in the succession of steps for dividing $2x^3 - 3x^2 + 4x - 1$ by $x^2 + 1$:

$$
\begin{array}{rl}
 & 2x - 3 \\
x^2 + 1 & \overline{)\,2x^3 - 3x^2 + 4x - 1} \\
 & \underline{2x^3 \phantom{{}-3x^2} + 2x \phantom{{}-1}} \\
 & - 3x^2 + 2x - 1 \\
 & \underline{- 3x^2 \phantom{{}+2x} - 3} \\
 & \phantom{2x^3 - 3x^2 +{}} 2x + 2
\end{array}
$$

The remainder in this case is $2x + 2$. Since the divisor $x^2 + 1$ is of the second degree, all remainders we get will be of the first degree or the zero degree. They will therefore be in the form $a + bx$, where a and b are real numbers. The remainder is of zero degree in the cases where b happens to be 0.

Now we sort all the polynomials in the ring into residue classes, by putting together in the same class all polynomials that have the same remainder when they are divided by $x^2 + 1$. We shall designate each class by a symbol of the form C_P, where P is a polynomial that belongs to the class. Thus each class has many names, and we recognize that two names stand for the same class when we see that the polynomials written as subscripts have the same remainder when we divide by $x^2 + 1$. For example, C_0, C_{x^2+1}, and $C_{x^4+3x^2+2}$ all stand for the same class, because the polynomials 0, $x^2 + 1$, and $x^4 + 3x^2 + 2$ all have the remainder 0 when they are divided by $x^2 + 1$. The polynomial x belongs to the class C_x. The polynomial a, where a is a real number, belongs to the class C_a. Since different real numbers have different remainders when they are divided by $x^2 + 1$ (each is its own remainder), no two real numbers belong to the same class.

We define operations of addition and multiplication of residue classes in the same way that we did for residue classes modulo 3 in Chapter III. To add two residue classes, pick a polynomial from each class, and add them. Then identify the class to which the sum of the polynomials belongs. That class will be the sum of the two classes. In the notation we are using, a member of each class is always put on display as a subscript in the name of the class. So we can add two classes by merely adding their subscripts. For example, $C_2 + C_3 = C_5$. $C_{2x} + C_7 = C_{2x+7}$.

To multiply two residue classes, we multiply a polynomial in one class by a polynomial in the other, and find the class that the product belongs to. Since a member of each class is put on display as a subscript in the name of the class, we can multiply two classes by merely multiplying their subscripts.

With these definitions of addition and multiplication, the system of residue classes has the structure of a ring and a number system. This follows from the fact that addition and multiplication in the system are commutative and associative, and multiplication is distributive with respect to addition. We can verify the commutativity of addition very easily. $C_P + C_Q = C_{P+Q}$, and $C_Q + C_P = C_{Q+P}$. But addition of polynomials is commutative, so $P + Q = Q + P$, and therefore $C_{P+Q} = C_{Q+P}$. The other properties are established by a similar argument. The zero element in the ring is C_0, because if C_P is the class of any polynomial P, $C_0 + C_P = C_{0+P} = C_P$.

Within this number system, there is a subset that is isomorphic to the real number system. This subset consists of all classes of the form C_a, where a is a real number (that is, a polynomial of zero degree). To prove that it is isomorphic to the real number system, we have to produce a one-to-one correspondence that preserves the operations of addition and multiplication. The mapping we use for this purpose is $a \leftrightarrow C_a$. The mapping is a one-to-one correspondence, because no two different real numbers belong to the same class. Under this mapping, if a and b are two real numbers, their images are C_a and C_b. The sum of the images is $C_a + C_b = C_{a+b}$, which is the image of the sum. The product of the images is $C_a \cdot C_b = C_{ab}$, which is the image of the product. Therefore the set of classes of the form C_a, where a is a real number, is isomorphic to the real number system. Since isomorphic systems are the same for all practical purposes, we can use the symbols of one system to represent the other. So now, instead of writing C_a for the class to which a real number a belongs, we shall simply write a. Thus 0 will stand for C_0, 1 will stand for C_1, and -1 will stand for C_{-1}.

Now that we have found the real numbers within our new system, we know that it is an extension of the real number system. Next we observe that it contains an element that is a solution to the equation $x^2 + 1 = 0$. To show that this is so, we must produce a class C_P which has the property that $(C_P) \cdot (C_P) + 1 = 0$. The class that has this property is the class C_x. In fact, $C_x \cdot C_x + 1 = C_x \cdot C_x + C_1 = C_{x^2} + C_1 = C_{x^2+1} = C_0 = 0$. Notice that we made use of the fact that when $x^2 + 1$ is divided by $x^2 + 1$, the remainder is 0, so that $C_{x^2+1} = C_0$.

The class C_x therefore has the same property as the complex number we called i before. In fact, since x^2 has the remainder -1 when we divide it by $x^2 + 1$, $C_x \cdot C_x = C_{x^2} = C_{-1} = -1$. Let us therefore use the symbol i to stand for C_x, since the rule $i^2 = -1$ holds for C_x as well.

Now only one more step remains to show that this system of residue classes is essentially the same as the complex number system we built up before. We show that it is isomorphic to the complex number system. First let us remind ourselves that every class in the system can be represented by the remainder that the members of the class have in common. Secondly, we recall that all remainders, when we divide by $x^2 + 1$, are first degree polynomials of the form $a + bx$ where a and b are real numbers. So every member of our new class of numbers can be represented in the form C_{a+bx}. Now, using the rules for addition and multiplication of residue classes, we see that C_{a+bx} can be written as $C_a + C_b \cdot C_x$. We have already agreed to write a instead of C_a, and b instead of C_b, because classes of this special form behave like real numbers. We have also agreed to write i for C_x, because the rule that $i^2 = -1$ correctly describes the behavior of this class. So we can write $C_a + C_b \cdot C_x$ in the form $a + bi$, with the understanding that a and b behave like real numbers and i has the property $i^2 = -1$. When we write it this way we see that the residue class C_{a+bx} is nothing but the complex number $a + bi$ in disguise. The system of residue classes modulo $(x^2 + 1)$ is therefore the field of complex numbers. This completes the third construction of the complex number system.

We have now seen the system of complex numbers in three forms of dress. We have seen it as a system of ordered pairs, as a system of two by two matrices, and as a system of residue classes in a ring of polynomials. In each of these systems, an individual complex number has a different appearance. But the structure of the three systems, in so far as their addition and multiplication tables are concerned, is the same. So we recognize them as merely different representations of one and the same number system, whose elements are usually represented in the convenient form $a + bi$.

No More Extensions Needed

We have expanded our number system four times. Each expansion to a more extensive system was made necessary by the fact that a certain type of equation could not be solved in the less extensive system. In the natural number system, we could not solve an equation like $x + 5 = 3$. So we constructed the system of integers, where all equations of the form $x + b = a$ have a solution. In the system of integers, we could not solve an equation like $2x = 3$. So we constructed the system of rational numbers, in which all equations of the form $ax = b$ have a solution, as long as a is not equal to 0. At this stage, we found that we could always solve equations of the first degree, whose general form is $ax + b = 0$. But we could not say the same for equations of the second degree. To be able to solve the equation $x^2 - 2 = 0$, we had to construct the real number system. To be able to solve the equation $x^2 + 1 = 0$, we had to construct the complex number system.

Now it is worth asking what equations we can solve in the complex number system. So far we have examined only equations of the first and second degree. If we try other algebraic equations of higher degree, will we find any that cannot be solved in the complex number system? Will it be necessary to expand the number system again and again, as we try more and more complicated algebraic equations? If you have had visions of an endless chain of extensions of the number system, you can dismiss them at once. As far as algebraic equations are concerned, we have reached the end of the road. The complex number system not only gives us a solution to the equation $x^2 + 1 = 0$. It gives us a solution for *every other algebraic equation* as well. To be more specific, it is known that every algebraic equation of the form

$$a_n x^n + a_{n-1} x^{n-1} \cdots a_1 x + a_0 = 0,$$

whose coefficients are complex numbers, has a solution in the complex number system. In fact it has as many solutions as the degree of the equation indicates. A first degree equation has one solution, a second degree equation has two solutions, a third degree equation has three solutions, and so on. The fact that algebraic equations can always be solved within the complex number system is known as the Fundamental Theorem of Algebra. It was first proved by the great German mathematician, Carl Friedrich Gauss (1777-1855).

The Old in the New

We have accomplished the purpose we aimed for at the beginning of the book. We have seen how, through successive extensions of the number system, mathematicians have eliminated its defects while losing none of its virtues. We have been introduced to a variety of mathematical structures, like groups, rings, fields, vector spaces, and topological spaces, that are being explored vigorously in mathematics today. We have found that, although their names are new, and at first sound unfamiliar, they are closely related to such familiar things as addition and multiplication of numbers, and collections of points on a line. Although the world of modern mathematics is a new world in many ways, it has never lost contact with the old world of number and space from which it has grown.

DO IT YOURSELF

1. Use the definition of multiplication of ordered pairs (see page 161) to find the following products:

   (2, 3) · (4, 1)   (2, 3) · (2, −3)   (2, 0) · (0, 3)

2. Write each of these ordered pairs in the form $a + bi$, making use of the convention that 1 stands for $(1, 0)$, and i stands for $(0, 1)$:

$(2, 3) \quad (2, -3) \quad (3, 0) \quad (0, 2)$

3. Verify by multiplication that $\frac{1}{2} - \frac{1}{2}i$ is the reciprocal of $1 + i$. (Remember that $i^2 = -1$)

4. Let 1 stand for the two by two matrix $\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$, and let i stand for the two by two matrix $\begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix}$. Write the two by two matrices that represent $2 + 3i$ and $2 - 3i$. Verify by matrix multiplication that $(2 + 3i) \cdot (2 - 3i) = 13 = \begin{vmatrix} 13 & 0 \\ 0 & 13 \end{vmatrix}$.

5. Use the rules for addition and multiplication of residue classes modulo $(x^2 + 1)$ to verify the following sums or products:

$C_{2x+1} + C_1 = C_0$. $\quad C_x \cdot C_{2x} = C_{-2}$.

$C_{x+1} \cdot C_{x+2} = C_{3x+1}$. $\quad C_{x^2} + C_3 = C_2$.

6. Multiply out, and express the answer in the form $a + bi$:

a) i^3	b) $(2 + i)^3$
c) i^4	d) $(3 + 4i)(5 - 2i)$
e) $(1 + i)^2$	f) $i(2 + 3i)$
g) $(2 - 3i)(2 + 3i)$	h) $(c + di)(e + fi)$

7. Prove that $(a + bi)(a - bi)$ is a real number.

8. Find the reciprocal of

a) i	b) $6 + 8i$
c) i^3	d) $1 + i$
e) $5 + 12i$	f) $\frac{1}{2} + \frac{\sqrt{3}}{2}i$

9. If $1 = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$ and $i = \begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix}$, write each of these complex numbers as a two by two matrix:

a) $1 + i$	b) $2 + 2i$	c) $2 + i$
d) $2 - i$	e) $3 + i$	f) $2 - 2i$
g) $-1 + i$	h) $3 - i$	i) $-1 - i$

10. Use the results of Ex. 9 to find these products by matrix multiplication; then check by the method of multiplication shown on p. 167:

a) $(1 + i)(3 + i)$ b) $(2 - 2i)(3 + i)$
c) $(2 - i)(2 + i)$ d) $(2 - 2i)(1 + i)$
e) $(-1 + i)(-1 - i)$ f) $(2 - i)(1 + i)$
g) $(3 + i)(3 - i)$ h) $(2 - 2i)(-1 - i)$

11. In any field, we can define division as follows: $A \div B = A \cdot \left(\frac{1}{B}\right)$. With the help of this definition, express in the form of $a + bi$:

a) $\frac{1 + i}{3 + 4i}$ b) $\frac{2 - i}{5 + 12i}$ c) $\frac{3 + i}{1 + i}$ d) $\frac{i}{1 - i}$

12. If $\bar{x} = a + bi$, $a - bi$ is called the conjugate of x, and is denoted by x. If x and y are complex numbers, prove:

a) $\overline{(\bar{x})} = x$.
b) $\bar{x} + \bar{y} = \overline{x + y}$.
c) $(\bar{x}) \cdot (\bar{y}) = \overline{(x \cdot y)}$.
d) $\bar{x} - \bar{y} = \overline{(x - y)}$.
e) $\frac{\bar{x}}{\bar{y}} = \overline{\left(\frac{x}{y}\right)}$.
f) $x + \bar{x}$ is a real number.
g) $\frac{x - \bar{x}}{i}$ is a real number.

Bibliography

For readers who wish to learn more about the subjects discussed in this book:

Elements of Algebra, by Howard Levi
Chelsea Publishing Co., New York
Higher Algebra for the Undergraduate, by Marie J. Weiss
John Wiley & Sons, New York
A Survey of Modern Algebra, by Garrett Birkhoff & Saunders MacLane, The Macmillian Company, New York

Summary of Basic Definitions

Zero element: A is a zero element for addition if $A+X=X+A=X$, for every X.

Unity element: U is a unity element for multiplication if $U\cdot X=X\cdot U=X$, for every X.

Identity element: E is an identity element for the operation $*$ if $E*X=X*E=X$, for every X.

Negative: A is the negative of B if $A+B=B+A=0$.

Reciprocal: A is the reciprocal of B if $A\cdot B=B\cdot A=1$.

Inverse: A is the inverse of B with respect to the operation $*$ whose identity element is E, if $A*B=B*A=E$.

Commutative: Addition is commutative if $x+y=y+x$. Multiplication is commutative if $x\cdot y=y\cdot x$.

Associative: Addition is associative if $(x+y)+z=x+(y+z)$. Multiplication is associative if $(x\cdot y)\cdot z=x\cdot(y\cdot z)$.

Distributive: Multiplication is distributive with respect to addition if $x\cdot(y+z)=x\cdot y+x\cdot z$.

Number system: A set of elements is a number system if it has two binary operations called addition and multiplication, each of which is commutative and associative, and if multiplication is distributive with respect to addition.

Group: A system is called a group if it has a binary operation that is associative, has an identity element for the operation, and has an inverse for every element. If the operation is commutative, the group is called abelian.

Ring: A system is called a ring if it has two associative binary operations called addition and multiplication, is an abelian group with respect to addition, and if the multiplication is distributive with respect to addition.

Field: A ring is called a field if it has a unity element for multiplication and contains a reciprocal for every element except 0.

Topological Space: A system of elements is called a topological space if a collection of its subsets is singled out to be called "open sets," and the collection has these properties: 1) The whole space and the empty set belong to the collection. 2) The union of any number of sets in the collection is also in the collection. 3) The intersection of any two sets in the collection is also in the collection.

Vector Space: A system is a vector space if it is an abelian group with respect to addition, is subject to a scalar multiplication by elements from an associated field of scalars, and if the scalar multiplication obeys these laws:

$$r \cdot (\vec{x} + \vec{y}) = r \cdot \vec{x} + r \cdot \vec{y}.$$

$$(r + s) \cdot \vec{x} = r \cdot \vec{x} + s \cdot \vec{x}$$

$$r \cdot (s \cdot \vec{x}) = (r \cdot s) \cdot \vec{x}$$

$$1 \cdot \vec{x} = \vec{x}.$$

Algebra: A system is an algebra if it is provided with binary operations of addition and multiplication, and a scalar multiplication, that make it both a vector space and a ring.

Geometry: A geometry is a study which identifies those figures in a space which are equivalent to each other under a group of transformations, and determines what properties equivalent figures have in common.

Do-It-Yourself Supplement

CHAPTER I

Ex. 1-4 are on pp. 33-4 of the text.

5. A multiplication operation is defined for the system of Ex. 2 by this table:

$\cdot$	a	b
a	a	a
b	a	b

a) Does this system have a unity element?

b) Show that multiplication is commutative in this system.

c) Show that multiplication is associative in this system. (Write the eight possible statements of the form $x \cdot (y \cdot z) = (x \cdot y) \cdot z$, and verify them from the table.)

d) Show that this multiplication is distributive with respect to the addition defined in Ex. 2. (Write the eight possible statements of the form $x \cdot (y + z) = x \cdot y + x \cdot z$, and verify them from the tables.)

e) Using the same tables, show that addition is not distributive with respect to multiplication. (Write the eight possible statements of the form $x + (y \cdot z) = (x + y) \cdot (x + z)$, and show that they are not all true.)

f) What conclusion follows from the results of 2b, 4a, b, and 5b, c, d?

6. Define addition and multiplication for the system of two elements 0 and 1 as follows:

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

a) Show that 0 is a zero element and that 1 is a unity element for the system.

b) Show that, with the operations defined by these tables, the system is a number system.

c) Show that this system is isomorphic to the system defined in Ex. 2 and 5.

7. Define addition and multiplication for the system of three elements a, b and c as follows:

$+$	a	b	c
a	a	b	c
b	b	c	a
c	c	a	b

$\cdot$	a	b	c
a	a	a	a
b	a	b	c
c	a	c	b

a) Show that a is a zero element and b is a unity element.
b) Show that the tables define a number system.

8. If a and b are both zero elements in a number system, prove that $a = b$.

9. If a and b are both unity elements in a number system, prove that $a = b$.

10. Show that $(20 \text{ av } 24) \text{ av } 36 \neq 20 \text{ av } (24 \text{ av } 36)$. (See p. 21.)

11. Define a binary operation D for the system of natural numbers (including 0) as follows:
If $a = b$, $aDb = 0$. If $a \neq b$, $aDb =$ the larger of the numbers minus the smaller one. (For example, $5D5 = 0$; $6D2 = 4$; $3D4 = 1$.)
a) Prove that the operation D is commutative.
b) Prove that the operation D is not associative.

12. Define a binary operation S for the system of natural numbers as follows: $aSb = 0$, if a is less than b; $aSb = 1$, if a is greater than b; and $aSa = a$.
a) Construct the table for the operation S.
b) Prove that the operation S is neither commutative nor associative.

13. Determine whether the binary operation $*$ defined for the system of natural numbers is commutative or associative if
a) $x * y = x + 3y$.
b) $x * y = x + y + 5$.

14. a) Use ordinary addition and multiplication as operations on the set $\{0, 3, 6, 9, \ldots\}$ containing all the natural numbers that are multiples of 3. Show that, with these operations, the system of multiples of 3 is a number system.
b) $k \longleftrightarrow 3k$ defines a one-to-one correspondence between the system of natural numbers and the system of multiples of 3. Show that this correspondence is *not* an isomorphism.

CHAPTER II

Ex. 1-3 are on p. 43 of the text.

4. Assign names to the subsets found in Ex. 3.
a) Construct a table for the union operation for this system of subsets.

b) Construct a table for the intersection operation.

5. Assign names to the subsets of $\{x\}$ as follows: $I = \{x\}$, $0 = \{\ \}$. Construct a table for the union operation and intersection operation for this system of subsets.

6. Let $I = \{a, b, c, d, e, f, g, h\}$. If X is a subset of I, let X' represent the complement of X in I.

Let $R = \{a, b, e, f\}$ $\quad S = \{a, b, g, h\}$ $\quad T = \{a, b, c, d\}$
$U = \{g, h\}$ $\quad V = \{a, b, c\}$ $\quad W = \{a, e, h\}$
$Z = \{b, d, g\}$

List the elements in each of these sets:

a) $R', S', T', U', V', W', Z'$.
b) $R \cup S, R \cup T, R \cup U, R \cup V, R \cup W$.
c) $R \cap S, R \cap T, R \cap U, R \cap V, R \cap W$.
d) $R \cup Z, S \cup T, S \cup U, S \cup V, S \cup W$.
e) $R \cap Z, S \cap T, S \cap U, S \cap V, S \cap W$.
f) $S \cup Z, T \cup U, T \cup V, T \cup W, T \cup Z$.
g) $S \cap Z, T \cap U, T \cap V, T \cap W, T \cap Z$.
h) $U \cup V, U \cup W, U \cup Z, V \cup W, V \cup Z, W \cup Z$.
i) $U \cap V, U \cap W, U \cap Z, V \cap W, V \cap Z, W \cap Z$.
j) $R \cup S', T \cup W', U \cup V', U \cup Z', W \cup Z'$.
k) $R' \cup S, T' \cup W, U' \cup V, U' \cup Z, W' \cup Z$.
l) $R \cap S', T \cap W', U \cap V', U \cap Z', W \cap Z'$.
m) $R' \cap S, T' \cap W, U' \cap V, U' \cap Z, W'' \cap Z$.

CHAPTER III

Ex. 1-6 are on pp. 71-2 of the text.

7. Add $(6 \sim 2) + (2 \sim 5)$; $(10 \sim 1) + (1 \sim 10)$; $(7 \sim 4) + (12 \sim 3)$.

8. Multiply $(6 \sim 2) \cdot (2 \sim 5)$; $(10 \sim 1) \cdot (1 \sim 10)$; $(7 \sim 4) \cdot (12 \sim 3)$.

9. Use the test for equality of integers to prove that $(5 \sim 2) = (7 \sim 4)$. (See p. 48.)

10. a) Add $(5 \sim 2) + (x \sim y)$.
b) Add $(7 \sim 4) + (x \sim y)$.
c) Use the text for equality of integers to show that the answers to a) and b) are equal.

11. a) Multiply $(5 \sim 2) \cdot (x \sim y)$.
b) Multiply $(7 \sim 4) \cdot (x \sim y)$.
c) Use the test for equality of integers to show that the answers to a) and b) are equal.

12. Prove $(a \sim b) + [(c \sim d) + (e \sim f)] = [(a \sim b) + (c \sim d)] + (e \sim f)$.

13. Prove $(a \sim b) \cdot [(c \sim d) \cdot (e \sim f)] = [(a \sim b) \cdot (c \sim d)] \cdot (e \sim f)$.

14. Prove $(a \sim b) \cdot [(c \sim d) + (e \sim f)] = (a \sim b) \cdot (c \sim d) + (a \sim b) \cdot (e \sim f)$.

15. Prove $(a + d \sim b + d) = (a \sim b)$.

16. If A is an integer, prove $(-1) \cdot A = -A$. (Hint: Let $A = (x \sim y)$. Then $-A = (y \sim x)$. $-1 = (0 \sim 1)$.)

17. Prove that in any ring with a unity element $(-1) \cdot A = -A$. (Hint: $(-1) + 1 = 0$. Multiply both sides by A on the right.)

18. Prove that if A and B are elements of a ring,
    a) $(-A) \cdot (B) = -(A \cdot B)$.
    b) $(A) \cdot (-B) = -(A \cdot B)$.
    c) $(-A) \cdot (-B) = A \cdot B$.
    d) $-(A + B) = (-A) + (-B)$.
    (Hint: Use the fact that $A + (-A) = 0$, and $B + (-B) = 0$.)

19. Define eight permutations on the letters A, B, C, D as follows:

I	P	Q	R	S	T	U	V
$A \to A$	$A \to B$	$A \to C$	$A \to D$	$A \to A$	$A \to C$	$A \to B$	$A \to D$
$B \to B$	$B \to C$	$B \to D$	$B \to A$	$B \to D$	$B \to B$	$B \to A$	$B \to C$
$C \to C$	$C \to D$	$C \to A$	$C \to B$	$C \to C$	$C \to A$	$C \to D$	$C \to B$
$D \to D$	$D \to A$	$D \to B$	$D \to C$	$D \to B$	$D \to D$	$D \to C$	$D \to A$

a) If consecutive vertices of a square are called A, B, C, D respectively, show that I, P, Q, R, S, T, U, V are the symmetries of the square.
b) Define * as in Ex. 4, and construct the multiplication table for this set of permutations.
c) Prove that they form a group with the operation *
d) Prove that I, P, Q, R form a subgroup.
e) Prove that I, S form a subgroup.
f) Find four more subgroups.

20. Use the associated remainders, 0, 1, 2, 3, as the names of the residue classes modulo 4. Construct addition and multiplication tables for these residue classes. What element is a zero divisor in this system?

21. Show that the set of all multiples of 4 is a subgroup of the group of integers with respect to the operation +. Show that it is a subring of the ring of integers. Show that it is an ideal of the ring of integers.

22. Every integer can be written in the form $3n + r$, where $r = 0, 1$, or 2, and n is an integer. $3n + r \to r$ defines a mapping of the ring of integers onto the residue class ring modulo 3. Show that this mapping is a ring homomorphism.

23. Show that the subgroup I, P, Q, R in Ex. 19 is isomorphic to the residue class group modulo 4 with + as group operation. (Ex. 20.)

24. Let K = the subgroup $\{I, P, Q, R\}$ in Ex. 19.
Denote the set $\{S * I, S * P, S * Q, S * R\}$ by $S * K$.
Denote the set $\{I * S, P * S, Q * S, R * S\}$ by $K * S$.
$S * K$ is called a left coset of K, and $K * S$ is called a right coset of K.

a) Show that in this case, $S * K = K * S$.

b) Show that $S * K = T * K = U * K = V * K$.

c) Denoting the coset obtained in b) by L, define the operation $*$ for the set $\{K, L\}$ as on p. 67, and construct the multiplication table for the set $\{K, L\}$ with this operation. Show that the set $\{K, L\}$ with the operation $*$ is a group (a quotient group). Show that it is isomorphic to the subgroup $\{I, S\}$.

CHAPTER IV

Ex. 1-4 are on p. 89 of the text.

5. Prove that multiplication of rational numbers is associative.

6. Prove that multiplication of rational numbers is distributive with respect to addition.

7. a) Add $\left(\frac{-3}{2}\right) + \left(\frac{x}{y}\right)$

b) Add $\left(\frac{-9}{6}\right) + \left(\frac{x}{y}\right)$

c) Use the test for equality of rational numbers to show that the answers to a) and b) are equal.

8. a) Add $\left(\frac{a}{b}\right) + \left(\frac{x}{y}\right)$

b) Add $\left(\frac{a \cdot c}{b \cdot c}\right) + \left(\frac{x}{y}\right)$

c) Use the test for equality of rational numbers to show that the answers to a) and b) are equal.

9. a) Multiply $\left(\frac{-3}{2}\right) \cdot \left(\frac{x}{y}\right)$

b) Multiply $\left(\frac{-9}{6}\right) \cdot \left(\frac{x}{y}\right)$

c) Use the test for equality of rational numbers to show that the answers to a) and b) are equal.

10. a) Multiply $\left(\frac{a}{b}\right) \cdot \left(\frac{x}{y}\right)$

b) Multiply $\left(\frac{a \cdot c}{b \cdot c}\right) \cdot \left(\frac{x}{y}\right)$

c) Use the test for equality of rational numbers to show that the answers to a) and b) are equal.

11. Prove that $\left(\frac{a}{c}\right) + \left(\frac{b}{c}\right) = \left(\frac{a+b}{c}\right)$

12. If $c \neq 0$, prove that $\left(\frac{a}{b}\right) = \left(\frac{a \cdot c}{b \cdot c}\right)$

13. Define "greater than" (denoted by $>$) for the rational number system as follows: $a > b$ if $a-b$ is positive. (See p. 54.) If a and b are rational numbers, and $a > b$, let $m = \frac{1}{2}(a+b)$. Prove that $a > m$, and $m > b$.

CHAPTER V

Ex. 1-7 are on p. 124 of the text.

8. The method used on page 107 for expressing a repeating decimal as a fraction can be used for finding the sum of some convergent series. For example, if $x = 1 + \frac{1}{2} + (\frac{1}{2})^2 + (\frac{1}{2})^3 + \ldots\ldots$, then $x = 1 + \frac{1}{2}[1 + \frac{1}{2} + (\frac{1}{2})^2 + \ldots]$ and $x = 1 + \frac{1}{2}x$. Solving this equation for x, we find that $x = 2$. Use this method to find the sum of each of these series:

a) $1 + \frac{1}{3} + \left(\frac{1}{3}\right)^2 + \left(\frac{1}{3}\right)^3 + \ldots\ldots$

b) $1 + \left(\frac{2}{3}\right) + \left(\frac{2}{3}\right)^2 + \left(\frac{2}{3}\right)^3 + \ldots\ldots$

c) $1 + \left(\frac{1}{4}\right) + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^3 + \ldots\ldots$

d) $2 + \frac{2}{3} + \frac{2}{3^2} + \frac{2}{3^3} + \ldots$

9. Which of these sets of real numbers are open? closed? neither open nor closed?

a) $\{1, 2, 3, 4, 5\}$
b) The set of all integers.
c) The set of all numbers x such that $2 < x < 3$.
d) The set of all numbers x such that $2 \leqq x < 3$.
e) The set of all numbers x such that $2 \leqq x \leqq 3$.
f) The set of numbers of the form $5 - \frac{1}{n}$, where n is a positive integer.

CHAPTER VI

Ex. 1-11 are on pp. 146-7 of the text.

12. Using A, B, C, D, as defined in Ex. 10,
a) find the products $A * B$ and $B * A$.
b) find the products $A * C$ and $C * A$.
c) find the products $A * D$ and $D * A$.

d) find the products $B * C$ and $C * B$.
e) find the products $B * D$ and $D * B$.
f) find the products $C * D$ and $D * C$.

CHAPTER VII

Ex. 1-4 are on p. 159 of the text.

5. Prove that
$$r\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} r & 0 \\ 0 & r \end{vmatrix} \cdot \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

6. Prove that
$$r\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = \begin{vmatrix} r & 0 & 0 \\ 0 & r & 0 \\ 0 & 0 & r \end{vmatrix} \cdot \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}$$

7. The transformation B is defined by:
$$x' = 2x + y$$
$$y' = 3x + 2y.$$

a) Find B^{-1}.
b) Write the matrices of B and B^{-1}.
c) If we denote these matrices by B and B^{-1} respectively, show that $B\,B^{-1} = B^{-1}\,B = I$.

8. If $A = \begin{vmatrix} r & 0 \\ 0 & s \end{vmatrix}$, and $B = \begin{vmatrix} x & 0 \\ 0 & y \end{vmatrix}$, prove that $AB = BA$.

9. If $A = \begin{vmatrix} r & 0 \\ 0 & r \\ 0 & r \end{vmatrix}$, and $B = \begin{vmatrix} x & y \\ z & w \end{vmatrix}$, prove that $AB = BA$.

10. If $A = \begin{vmatrix} r & 0 \\ 0 & s \end{vmatrix}$, and $B = \begin{vmatrix} x & y \\ z & w \end{vmatrix}$, find AB and BA.

11. If $A = \begin{vmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{vmatrix}$, and $B = \begin{vmatrix} x & 0 & 0 \\ 0 & y & 0 \\ 0 & 0 & z \end{vmatrix}$, prove that $AB = BA$.

12. If $A = \begin{vmatrix} c & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & c \end{vmatrix}$, and $B = \begin{vmatrix} x & y & z \\ u & v & w \\ r & s & t \end{vmatrix}$, prove that $AB = BA$.

13. If $A = \begin{vmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{vmatrix}$, and $B = \begin{vmatrix} x & y & z \\ u & v & w \\ r & s & t \end{vmatrix}$, find AB and BA.

Index

Abelian Group, 60, 66-7, 84, 130, 141, 150, 162
Absolute value, 114
Addition, 17, 30, 48, 64, 75-6, 104, 128, 130, 149, 174
Adherence, point of, 117-8
Affine geometry, 146
Affine group, 144
Affine transformation, 144
Algebra, 156, 163
Algebraic equations, 90-2
Arabic numerals, 25
Associative law, 20, 22, 40, 84, 130, 132, 141, 150-1, 152, 155, 161, 175

Binary operation, 19
Boolean algebra, 43

Cancellation law, 69
Cardinal number, 14-6
Closed interval, 118
Closet set, 118
Cluster point, 117
Commutative group, 60, 107
Commutative law, 19, 22, 40, 49, 60, 77, 105, 130, 150-1, 161, 169, 175
Complement, 42, 118
Complex numbers, 160-78
Congruent, 145
Convergent series, 111
Co-ordinates, 126
Coset, 66, 68

Decimal fractions, 95
De Morgan's rule, 42
Descartes, 168
Determinant, 144
Difference, 44
Dimension, 134-5
Distributive law, 23, 40, 107, 132, 151, 155, 162, 175
Division, 19, 73, 80

Empty set, 30, 118-9
Euclidean group, 144-6

Factorable integers, 87
Field, 84-7, 107, 163, 172
Finite field, 84-7
Fraction, 73, 75
Fundamental theorem of algebra, 178

Galois, 60
Gauss, 178
Geometry, 145
Group, 54-5, 59, 83-4, 140-5
Group of similitudes, 144

Hebrew numerals, 25
Homeomorphism, 123
Homomorphism, 70

Ideal, 64, 68, 87
Identity element, 29, 55, 84, 141
Image, 14
Imaginary numbers, 168
Imaginary unit, 166-7
Infinite decimal, 96-9, 103-5, 109
Infinite sequence, 97, 110, 113
Infinite series, 109, 113
Integers, 47-54, 81
Intersection, 38-9
Inverse, 55, 84, 141, 143
Irrational numbers, 93, 102

Isomorphism, 26, 52, 70, 81, 105, 123, 164, 170, 175-6

Larger number, 25, 54, 83
Limit, 98, 110
Limit point, 115-6
Linear group, 141-4
Linear mapping, 139, 148

Many-to-one mapping, 15
Mapping, 14
Matrices, 149, 168-71
Matrix equation, 158
Multiplication, 21, 30, 48, 65, 75-6, 104, 153, 174

Natural numbers, 19, 32-3, 52
Negative, 50-1, 55, 66, 78-9, 107, 130, 150-1
Negative integer, 53
Negative rational number, 82
Neighborhood, 115-6, 119-20
Nest of intervals, 98-102, 108-9, 111
Nilpotent, 156
Normal subgroup, 67
Number system, 24, 40, 49, 64, 77, 103, 162, 175

One-to-one correspondence, 14-5
One-to-one mapping, 15
Open interval, 119
Open set, 118
Ordered pair, 18, 75, 126, 157, 160

Partial sum, 110
Peano's axioms, 32-3
Place value, 27
Points in a plane, 126, 160
Point of adherence, 117-8
Points on a line, 29, 54, 82, 94, 105, 108, 115
Polynomials, 134, 172
Positive integer, 52
Positive rational number, 82
Prime integers, 87
Pythagoras, 94

Quotient, 73, 75
Quotient group, 66
Quotient ring, 67

Rational numbers, 76-80, 105
Real numbers, 103, 164-5, 168, 175
Reciprocal, 79, 84, 107, 163
Rectangular matrices, 157
Reflection, 137
Repeating decimals, 105
Residue classes, 62-5, 85-7, 171-6
Resultant, 129
Reversible mappings, 139
Ring, 60, 66, 74, 83-4, 107, 155, 162, 172-3, 175
Roman numerals, 25

Scalar, 132
Scalar multiplication, 132-3, 149
Set, 36
Similar, 145
Similitude, 144
Standard set, 16
Subgroup, 63, 67
Subring, 63
Subset, 36
Subtraction, 44, 52
Symmetries of a triangle, 55-9

Topological space, 115, 120
Transformation, 137
Translation, 139

Union, 37-8
Unit ideal, 88
Unity element, 28-9, 41, 50, 64, 78, 84, 107, 133, 151, 153, 155, 163, 173

Vector, 132
Vector space, 133-6, 149, 160, 162, 172

Zero, 28, 61, 74
Zero divisor, 69, 87, 156
Zero element, 28-9, 41, 50, 65, 78, 84, 107, 130, 150-1, 172, 175
Zero ideal, 88

What are "real" numbers?

What are "imaginary" numbers?

Can there be a number system *without* numbers?

We learn to count almost as soon as we learn to speak. What we rarely learn are the *hows* and *whys* of our workday numbers. Irving A[illegible] book is the first to explain the familiar 1, 2, 3, and [illegible]f hiding the basic and profound concepts that are found there. He traces advanced theories from their beginning in grammar-school arithmetic, and helps the layman to discover the meaning of the revolutionary developments in the world of modern mathematics.

IRVING ADLER is an Instructor in Mathematics in the School of General Studies at Columbia University and the author of more than a dozen books, including **How Life Began**, **Magic House of Numbers**, and **The Stars**. The hardcover editions of his books are published by The John Day Company.

PUBLISHED BY THE NEW AMERICAN LIBRARY